ARO-D Report 63-2

PROCEEDINGS OF THE EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

Sponsored by
The Army Mathematics Steering Committee
on Behalf of

THE OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE Dec. 1963	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
Proceedings of the Eighth Conference on the Design of Experiments in Army Research Developments and Testing

5. FUNDING NUMBERS

6. AUTHOR(S)
Not Available

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Army Mathematics Steering Committee

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO-OORR 63-2

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This is a Technical report resulting from the Proceedings of the Eighth Conference on the Design of Experiments in Army Research Developments and Testing.

14. SUBJECT TERMS

15. NUMBER OF PAGES
585

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 63-2
December 1963

PROCEEDINGS OF THE EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by the Army Mathematics Steering Committee

conducted at

Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D. C.
24-26 October 1962

U. S. Army Research Office-Durham
Box CM; Duke Station
Durham, North Carolina

TABLE OF CONTENTS

---

* This paper was presented at the Conference. It does not appear in these proceedings.

** This paper was not presented at the Conference. It does not appear in this technical manual.

TABLE OF CONTENTS (contd.) Page

* This paper was presented at the Conference. It does not appear in these proceedings.

TABLE OF CONTENTS (contd) Page

* This paper was presented at the Conference. It does not appear in these proceedings.

FOREWORD

The Walter Reed Army Institute of Research, Walter Reed Army Medical Center served as the host for the Eighth Conference on the Design of Experiments in Army Research, Development and Testing. This Washington, D. C. Conference was held 24-26 October 1962. Colonel Conn L. Milburn, Jr., Director of WRAIR, issued the following letter to those attending this scientific meeting:

"The staff and faculty of the Institute feel privileged and honored that Walter Reed Army Institute of Research has been selected as the place for this Eighth Conference on the Design of Experiments in Army Research, Development and Testing. To each of you we extend a most cordial welcome.

It is a pleasure to have you with us and we sincerely hope that your stay here will be both enjoyable and professionally rewarding."

The Army Mathematics Steering Committee takes this opportunity to thank Colonel Milburn for his welcoming remarks and for the use of the excellent facilities under his command. The thanks of the Committee are also due to the two Local Chairmen, Lt. Col. Stefano Vivona and Major Paul J. Wentworth, appointed by him, for their very efficient handling of all local arrangements.

The following information about the host installation was extracted, with minor alterations, from the Medical Annals of the District of Columbia, Vol XXX, No. 11, November 1961.

"The Walter Reed Army Institute of Research (WRAIR) was founded in 1893 as the Army Medical School, the first school of preventive medicine in the United States. The school was established by Surgeon General George M. Sternberg, one of the great military men of science, thus bringing to fruition an idea first proposed by Surgeon General William A. Hammond in 1862. On General Sternberg's recommendation the War Department issued General Orders No. 51, dated 24 June 1893, founding the school.

Captain (later Major) Walter Reed, in whose honor the Army Medical Center is named, was a member of the first faculty. Captain Reed was Professor of Clinical and Sanitary Microscopy and Director of the Pathological Laboratory. He was also the first Secretary of the Faculty.

The Army Medical School marked its fifthieth anniversary on 18 December 1943 with graduation exercises for the class completing courses in military and tropical medicine. The class of 124 included medical officers from the armies of the United States, Canada, and Peru, and officials of state health departments. An honored guest at the exercises was Colonel Deane C. Howard, retired, who held the highest rating in the first class graduated from the Army Medical School in 1894.

Since World War II the School has changed its name several times. The name, Army Medical Department Research and Graduate School, was adopted in 1947. This was changed to The Army Medical Service Graduate School in 1950 and finally, in 1955, the present name, The Walter Reed Army Institute of Research, was adopted.

The basic mission of the Walter Reed Army Institute of Research is 'To provide the medical research and professional graduate training required by the Army to fulfill its role in the national defense.' Actually, this broad mission is divided into 4 parts: (1) serving as a research and development center, (2) providing education and training to officers of the Army Medical Service (also to some officers from other branches of the Armed Forces and from other countries and some civilians), (3) serving as a central reference laboratory for the Army Medical Service and, in a number of fields, for the Navy, Air Force and Veterans Administration, and (4) manufacturing more than 100 biological products, not obtainable through commercial sources, which are supplied to all the Armed Forces and the Veterans Administration.

The first class began on 1 November 1893 with 5 students officially enrolled. From this small beginning the original Army Medical School has grown into the present great Institute of Research whose activities cover all parts of the world."

At the Eight Conference, invited addresses were delivered by Drs. Robert P. Abelson, Herbert C. Batson, Herman Chernoff, Egon S. Pearson, and Marvin A. Schneiderman. Decisions under uncertainty, bio-assay, optimal designs, assessing optimal performance of weapons, screening theory were, respectively, the areas discussed by these five specialists. Dr. Harold F. Dorn served as Chairman for the Panel Discussion on Diet and Heart Disease. Dr. George V. Mann initiated the Panel Discussion by presenting data and known facts about diet and heart diseases. Mr. Jerome Cornfield commented on the material presented by Mann; then he and the Chairman led the discission and answered

questions by members of the audience. Miss Beatrice Orleans showed and discussed a film on the Design of Experiments. Two other important parts of the conference were the three papers discussed in the Clinical Sessions, and the twenty-six papers delivered in the Technical Sessions.

The present volume of the Proceedings contains twenty-eight papers which were presented at this meeting. The Army Mathematics Steering Committee, the sponsor of this series of conferences, has asked that these articles on modern statistical principles in the Design of Experiments, as well as certain applications of these ideas, be made available in the form of these Proceedings.

The Eighth Conference was attended by registrants and participants from over 80 different organizations. Speakers and panelists came from Booz-Allen Applied Research, Inc.; Bureau of Ships; C-E-I-R, Inc.; Ford Motor Co.; Iowa State University of Science and Technology; National Bureau of Standards; National Cancer Institute, NIH; National Heart Institute, NIH; North Carolina State College; Operations Research, Inc.; Princeton University; Research Analysis Corporation; Research Triangle Institute; Stanford University; University College, London; University of Illinois, College of Medicine; University of Michigan Institute of Science and Technology; University of North Carolina; University of Wisconsin, U. S. Army Mathematics Research Center; Vanderbilt University; Virginia Polytechnic Institute; Yale University, and thirteen Army Facilities.

Before closing, the Chairman wishes to express his sincere thanks to his Advisory Committee: F. G. Dressel (Secretary), Fred Frishman, B. G. Greenberg, Frank E. Grubbs, Boyd Harshbarger, H. L Lucas, Jr., Clifford J. Maloney, Lt. Col. Stephano Vivona, and Marvin Zelen for their suggestions and assistance in selecting the invited speakers and formalizing the plans for this conference. The Chairman is especially grateful to Dr. Dressel for coordinating the Conference program and seeing these Proceedings through publication.

S. S. Wilks
Professor of Mathematics
Princeton University

EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

24-26 October 1962

Walter Reed Army Institute of Research

Wednesday, 24 October

0900-0930 REGISTRATION
Lobby of Sternberg Auditorium (WRAIR)

0930-1215 GENERAL SESSION I
Sternberg Auditorium

Calling of Conference to Order
Lt. Colonel Stefano Vivona, Local Chairman

Welcome to Walter Reed Army Medical Center
Major General A. L. Tynes, Commanding

Welcome to Walter Reed Army Institute of Research
Colonel Conn L. Milburn, Jr., Commanding

Announcements
Major Paul J. Wentworth, Chairman on Local Arrangements

Chairman: Dr. Churchill Eisenhart, Statistical Engineering Laboratory, National Bureau of Standards

A Statistician's Place in Assessing the Likely Operational Performance of Army Weapons and Equipment
Professor Egon S. Pearson, University College, London

A General Survey of Screening Theory
Dr. Marvin A. Schneiderman, National Cancer Institute, National Institutes of Health

1220-1320 LUNCH - Ballroom, Officers' Open Mess, WRAMC

Technical Sessions I and II as well as Clinical Session A will start at 1330. After the coffee break Technical Session III and IV will convene and run from 1510-1700. Starting at 1700 there will be a Social Hour. This will provide an opportunity for old friends to get together and for new friends to get acquainted. We hope that everyone at the conference will be able to stay for this phase of the meeting.

1330-1440 TECHNICAL SESSION I - Sternberg Auditorium

Chairman: Sidney Sobelman, U.S. Army Munitions Command, Picatinny Arsenal

Estimation of Service Life from Fatigue Testing Results on Full Scale Specimens
John P. Purtell, Research and Engineering Division, Watervliet Arsenal.

The Simulated versus National Environment in Military Testing and Operations
C. Bruce Lee, Systems Analysis Section, Advanced Design Branch, Research & Engineering Directorate, Ordnance Tank-Automotive Command, Center Line, Michigan

1330-1440 TECHNICAL SESSION II - Room 358

Chairman: A.C. Cohn, Jr., The University of Georgia

Application of 2^{8-4} Fractional Factorials in Screening of Variables Affecting the Performance of Dry Process Zinc Battery Electrodes
Nicholas T. Wilburn, U.S. Army Electronics R & D Laboratory, Fort Monmouth, New Jersey

Applications of the Calculus for Factorial Arrangements
B. Kurkjian, Diamond Ordnance Fuze Laboratories
M. Zelen, U.S. Army Mathematics Research Center, University of Wisconsin

1330-1440 CLINICAL SESSION A - Room 341

Chairman: Frank E. Grubbs, Ballistics Research Laboratories

Panelists: W.T. Federer, U.S. Army Mathematics Research Center, The University of Wisconsin
B.G. Greenberg, The University of North Carolina
H.O. Hartley, Iowa State University of Science & Technology
H.L. Lucas, North Carolina State College
M.A. Schneiderman, National Institutes of Health

Statistical Procedures for the Evaluation of Thrust Curves
Paul C. Cox, White Sand Missile Range, New Mexico

1440-1510 COFFEE BREAK
Lobby Sternberg Auditorium

1510-1700 TECHNICAL SESSION III - Room 341

Chairman: Gerhard J. Isaac, U.S. Army Medical Research and Nutrition Laboratory, Denver, Colorado

The Independent Action Theory of Mortality as Tested at Fort Detrick
Francis M. Wadley, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Trial and Station Variability in P^{32} for Tripartite Collaborations
Walter D. Foster, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Design and Analysis of Entomological Field Experiments
William A. Brown, Dugway Proving Ground, Dugway, Utah
Scott A. Krane, C-E-I-R, Inc., Dugway Field Office, Dugway, Utah

1510-1700 TECHNICAL SESSION IV - Sternberg Auditorium

Chairman: Vaughn LeMaster, Ammunition Procurement and Supply Agency

Comparison of Two Approaches to Obtaining a Transformation Matrix Effecting a Fit to a Factor Solution Obtained in a Different Sample
Cecil D. Johnson, U.S. Army Personnel Research Office

Some Least-Squares Transformations of Regression Estimators of Orthogonal Factors
Emil F. Heermann, U.S. Army Personnel Research Office

A Reliability Test Method for "One-Shot" Items
H. J. Langlie, Aeronutronics Division, Ford Motor Company

1700-1900 SOCIAL - Ballroom, Officers' Open Mess, WRAMC

Thursday, 25 October

Technical Session V and VI will be held from 0800-0920. Clinical Session B and Technical Session VII will be in order from 0950-1120. The first paper in Clinical Session B carries a classification of SECRET and the second paper a classification of CONFIDENTIAL. General Session 2 will start at 1130. After lunch Technical Session VII and IX will convene at 1330 and end at 1440. The Panel Discussion on Diet and Heart Disease is scheduled to start at 1510. A film entitled "The Design of Experiments" will be shown following the panel discussion.

0800-0920 TECHNICAL SESSION V - Sternberg Auditorium

Chairman: Ralph Brown, U.S. Army Materiel Command
Frankford Arsenal

Investigation in Temperature Control of Hydraulic Systems in Rough Terrain Fork Trucks
Irving Tarlow, Quartermaster Research and Engineering Command, Natick, Massachusetts

Precision of Simultaneous Measurement Procedures
W.A. Thompson, Jr., University of Delaware and National Bureau of Standards

0800-0920 TECHNICAL SESSION VI - Room 341

Chairman: B.G. Greenberg, The University of North Carolina

The Ultraviolet Microscopy of Tissues
George I. Lavin, Ballistic Research Larobatories, Aberdeen Proving Ground, Maryland

Redundancies in Human Biomechanics and their Application in Assessing Military Man-Task Disability Performance Resulting from Ballistic Agents
William H. Kirby, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

0920-0950 COFFEE BREAK - Lobby Sternberg Auditorium

0950-1120 TECHNICAL SESSION VII - Room 358

Chairman: Paul C. Cox, White Sands Missile Range

Half-Normal Plots for Multi-Level Factorial Experiments
Scott A. Krane, C-E-I-R, Inc., Dugway Field Office, Dugway Proving Ground, Dugway, Utah

Proportional Frequency Designs
Sidney Addelmann, Statistics Research Division, Research Triangle Institute

0950-1120 CLINICAL SESSION B - Room 341

Chairman: Clifford J. Maloney, National Institutes of Health

Panelists: O.P. Bruno, Ballistics Research Laboratories
Boyd Harshbarger, Virginia Polytechnic Institute
S.S. Wilks, Princeton University
Marvin Zelen, U.S. Army Mathematics Research Center, The University of Wisconsin

Security Classification: SECRET

Design of an Experiment to Evaluate Aerosol and Storage Characteristics of a Viral Slurry
Samuel N. Metcalfe, U.S. Army Biological Laboratories, Fort Detrick, Maryland
Bertram W. Haines, U.S. Army Biological Laboratories, Fort Detrick, Maryland

Security Classification: CONFIDENTIAL

Sensitivity Analysis of Outputs from a Computer Simulation Model of a Chemical Weapons System
Reynold Greenstone, Operations Research, Inc., Silver Spring, Maryland
Ira A. DeArmon, U.S. Army CBR Agency, Operations Research Group, Army Chemical Center, Maryland

1130-1230 GENERAL SESSION 2 - Sternberg Auditorium

Chairman: Fred Frishman, Army Research Office, Office, Chief of R & D

Optimal Design of Experiments
Professor Herman Chernoff, Stanford University

1230-1330 LUNCH - Ballroom, Officers' Open Mess, WRAMC

1330-1440 TECHNICAL SESSION VIII - Room 341

Chairman: Ira A. DeArmon, Jr., Operations Research Group, Army Chemical Corps

Vibration Experiments
F. Pradko, Dynamic Simulations Laboratory, U.S. Army Ordnance Tank-Automotive Command

Size Effects in the Measurement of Soil Strength Parameters
B. Hanamoto, Land Locomotion Lab, Research Division, Research and Engineering Directorate, ATAC
Emil H. Jebe, Institute of Science & Technology, The University of Michigan

1330-1440 TECHNICAL SESSION IX - Sternberg Auditorium

Chairman: Edwin Cox, Agriculture Research Service,
Beltsville, Maryland

Effectiveness of Certain Experimental Plans Utilized in Sensory Evaluations

J. Wayne Hamman and Jan Eindhoven, Quartermaster Food and Container Institute for the Armed Forces, Chicago, Illinois

An Evaluation of Radiation-Processed Foods for the Military Rations

Donald M. Boyd, Research Analysis Corporation

1440-1510 COFFEE BREAK - Lobby Sternberg Auditorium

1510-1700 GENERAL SESSION 3 - Sternberg Auditorium

Panel Discussion on Diet and Heart Disease

Chairman: Dr. Harold F. Dorn, Biometrics Research Branch, National Heart Institute

Panelists: Dr. George V. Mann, School of Medicine, Vanderbilt University

Mr. Jerome Cornfield, Biometrics Research Branch, National Heart Institute

1700-1730 SPECIAL SESSION - Sternberg Auditorium

Chairman: Beatrice Orleans, Bureau of Ships

Film on the Design of Experiments

Friday, 26 October

Technical Sessions X and XI run from 0800-1000. General Session 4 will start at 1030 and end at 1230. After the lunch hour your host has planned for you an interesting and informative tour.

0800-1000 TECHNICAL SESSION X - Room 341

Chairman: Earl Atwood, Walter Reed Army Institute of Research

Some Consequences of Some Assumptions with Respect to the Physical Decay of a Chamber Aerosol Cloud
Theodore W. Horner, Booz-Allen Applied Research, Inc., Bethesda, Maryland

The Role of Intuition in the Scientific Method
Nicholas M. Smith, Research Analysis Corporation

How to Design War Games to Answer Research Questions
W. L. Pierce, Research Analysis Corporation

0800-1000 TECHNICAL SESSION XI - Sternberg Auditorium

Chairman: H. L. Lucas, North Carolina State College

Evaluation of Performance Reliability
Seymour K. Einbinder, Picatinny Arsenal
Ingram Olkin, Stanford University

Evaluation of Various Laboratory Methods for Determining Reliability
A. Bulfinch, Picatinny Arsenal

Confidence Intervals for Systems Reliability
Jasper Dowling, Picatinny Arsenal

1000-1030 COFFEE BREAK - Lobby Sternberg Auditorium

1030-1230 GENERAL SESSION 4 - Sternberg Auditorium

Chairman: Dr. S. S. Wilks, Princeton University

An Experimental Design for Decisions under Uncertainty
Dr. Robert P. Abelson, Yale University

Bio-Assay
Dr. Herbert C. Batson, University of Illinois, College of Medicine

1230-1330 LUNCH - Ballroom, Officers' Open Mess, WRAMC

1330-1500 TOUR - Assemble in Sternberg Auditorium

A STATISTICIAN'S PLACE IN ASSESSING THE LIKELY OPERATIONAL PERFORMANCE OF ARMY WEAPONS AND EQUIPMENT

E. S. Pearson
University College, London, England

THE BACKGROUND OF THIS PAPER. It has been a special honour to receive an invitation from the organising committee of this Conference to make the journey from England and to address you today. In thinking how I could best repay the compliment, it seemed to me that I should look for a subject in illustrating which I could draw on my own particular experiences, gained in working for the British armed services both during and since the second world war. From 1939 to 1946 I was attached with a number of members of the University College, London Statistics Department, to the British Ordnance Board. This is an organisation of some historic interest for I believe its foundation can be traced back to an appointment made in 1414, the year before the Battle of Agincourt! It is now concerned with certain aspects of the development and acceptance of weapons for both the Army, the Navy and the Air Force. Then, for some years after the war, I was a member of the Ordnance Board Anti-aircraft Lethality Committee and very recently I have been pulled back to be chairman of an advisory committee concerned with the general problem of assessment in connection with army weapons and equipment.

My main experience was with the subject which has been described as terminal ballistics and in particular with the lethal effectiveness of anti-aircraft fire. We were concerned also with field artillery fire and with the medium and small bombs of those days, in so far as fragmentation of the casing rather than blast played an important part in their effectiveness. It is of course true that the weapons and the army requirements of 15-20 years ago have been to a large extent out-dated, but if I make my main topic today a piece of historical recording, it is because I believe that a number of general principles and lessons emerge from such a study which are still relevant to the practice of experimentation and analysis in Army Research today.

It seemed to me that there were two advantages in taking illustrations from World War II experience. In the first place I could speak of matters about which I had the 'feel' from first hand knowledge and so perhaps could be more interesting as well as convincing in any arguments put forward. Secondly, it was easier to be factual without running into the danger of using classified material. What I shall try to do, therefore, is to give you

first some account of the difficulties with which we were faced in the years 1939-45 in constructing a model which could be used to help determine how to improve the effectiveness of anti-aircraft fire. In describing this problem, it should be possible to indicate a number of lessons which are still relevant in a much wider field. There are also many points of difference which it will be instructive to emphasise.

THE STATISTICIAN'S PLACE. I should perhaps confess straight away that I shall say very little about statistics or about what is commonly thought of as the design of experiments. To this extent you may think that the leading phrase in the title of this paper is misleading, unless you interpret the words in the personal sense as referring to the statistician who is giving this address! But there is, I think, a point here which I should like to make. At the fourth of this series of Conferences, held in 1958, Dr. A. W. Kimball read a paper with the title: "Errors of the 3rd kind in statistical consulting"; in this he discussed and illustrated the fault of giving a perfectly sound statistical answer to a problem which is not the real one needing solution.

Many of us are I think conscious of what might perhaps be called an error of a 4th kind; that which the statistician makes when he allows his interest in the statistical elements of a problem and its potential for statistical elegance and sophistication to obscure what should be his prime objective, the solution of the real matter at issue. The fault is not so much that wrong statistical methods are used (Kimball's 3rd kind of error) but that the situation does not justify the use of any refined statistical methods at all until the outstanding problem has been solved of obtaining data which are both relevant and reliable. The statistician, indeed, is called upon to be a scientist in the fullest sense of that term--to apply scientific method, not merely statistical techniques, to the job on hand.

When he has completed some piece of mathematical or arithmetical analysis, he needs to ask himself searchingly: does this answer make sense? I can recall, as no doubt some of you can too, war-time reports which appeared both in my country and in yours, containing a pretty piece of algebraic development or some standard analysis of variance, the conclusions from which obviously did not make sense. Perhaps such reports from youthful enthusiasts would never have appeared but for the inevitable shortage of experienced and critical supervision in rapidly expanding organisations. They are likely, however, to discourage the idea that mathematics or statistics were of value in problems of weapon development and testing, because the experienced non-statistical layman, the military or naval technical officer who had the feel of the problems, could see at

once that the data would not bear the confident interpretation which was often placed on them.

Certainly in my own experience at the Ordnance Board it was the physical difficulty in securing meaningful experimental data which had always to be faced. There was very little opportunity for design as it is understood in agricultural or biological trials. There was no paramount function for the application of advanced statistics--we used to say that the only statistical tools which were needed were the normal distribution in 1, 2 and 3 dimensions, the Poisson and the bionomial. But it is true to say that the statistician's training, with the understanding which should follow of the meaning of variation and correlation, of randomness and probability, with its emphasis on the importance of adopting a critical outlook on assumptions--all this is likely to provide an excellent preparation for the kind of work we are discussing, but on one essential condition--that the training has been carried out in conjunction with practical application to data analysis. The trend in the teaching of mathematical statistics at our universities today is often increasingly away from any real application to data.

There is another point which I think is worth emphasising. One of the surest ways to cure the statistician from any tendency to over-sophistication is to arrange that he is present at experiments or trials, the data from which he is to use. In this respect we were lucky in England; we attended firing trials on the Shoeburyness Ranges, we were hot on the scene after bombs had been dropped on parked aircraft, trucks and wooden dummies in slit trenches on a special bombing range in the New Forest, and--as a wartime experience--we might happen to be present at a gun-site when German aircraft were the target. Under such conditions it is easier to come to grips with the meaning and limitations of data.

THE ANTI-AIRCRAFT PROBLEM. First let me try to put this problem into its setting of 20 or more years ago. As far as the Ordnance Board group was concerned, we had not to consider the problems of the deployment of guns, of the acquisition of targets, of the handling of mass attacks or other important tactical matters. These were questions for the Anti-aircraft Command and its Operational Research Section which was formed in the summer of 1940. Our work was closely related to the question of design, to understand more clearly the individual relationship between predictor, gun, shell, fuze and enemy target in order to advise what improvements were possible and likely to be worthwhile.

In this field of research where the terminal action in which one is interested may be taking place several thousand feet above ground, no overall experiment bringing in all the factors concerned is conceivable; the reasons for this are so obvious that I do not need to list them. As a consequence, it is absolutely essential to construct a mathematical model of the terminal engagement, and then to consider how the parameters of this model may best be estimated. As in so many other problems of military science, the model even if necessarily simplified, serves as an essential means of defining the relationships of the situation, showing how research investigation can be broken into separate pieces and emphasising at what points our lack of sure information is greatest and most hampering.

Let me now outline the problem and its solution in some detail, first describing the mathematical model and then discussing the three main headings under which gaps in knowledge had to be filled, namely:

(i) positioning errors (until the introduction of the proximity fuze in 1943-44 it was easy to combine the error of the time fuze with the predictor, gun-laying and ballistic errors);

(ii) fragmentation characteristics of the shell;

(iii) target vulnerability.

The difficulties which had to be overcome, largely through ignorance of physical properties in this hitherto unexplored field, are I think sufficiently instructive to be worth including as part of the story. Much the same problems were I know faced later on (building perhaps on our experience) in Section T of the Applied Physics Laboratory at Silver Spring and the associated Proving Ground near Albuquerque, where research and trial work was carried out for the U. S. Navy. I did not myself have any direct contact with U. S. Army investigations.

THE MATHEMATICAL MODEL. The first simplified model which was used involved:

(a) A three-dimensional normal distribution of positioning errors about the target, with a major axis along the shell trajectory and the standard errors in directions perpendicular to this axis equal, i. e., the density contours were taken to be ellipsoids with circular cross sections in planes perpendicular to the principal axis.

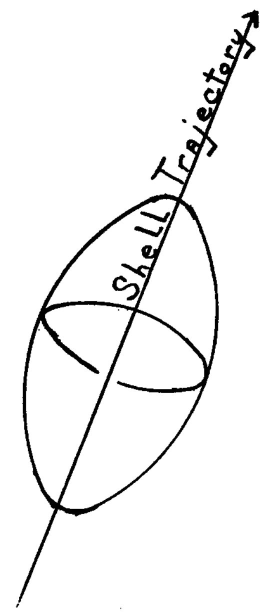


Figure 1

(b) A main fragment zone lying between two cones whose axis was that of the shell axis and the trajectory at time of burst, and a small sudsidiary nose cone.

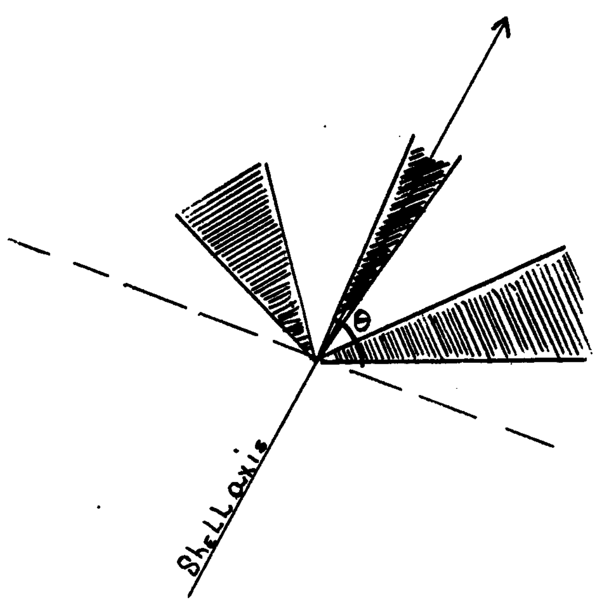


Figure 2

The density of fragments within the main zone was not of course uniform, though it might be treated as such for a first approximation. For any zone within which the average density of fragments of a given penetrating power could be regarded as constant, the probability distribution of strikes was taken as Poisson.

(c) For the aircraft, we first used what was termed an 'equivalent vulnerable target' represented by a sphere of a few feet in radius such that its 'perforation' by at least one 'lethal' shell fragment would result in a kill. Later, this representation had to be treated in more detail.

This simple model based on the trivariate normal and the Poisson distributions, with bounding surfaces consisting of ellipsoids, cones and spheres was amenable to computation, provided that meaningful numerical values for the various parameters could be estimated. But the task of filling in these unknown elements was immense and for a time the more we learnt, the more we realised our ignorance. Consider then some of the gaps to be filled.

THE POSITIONING ERRORS. The original data were collected from Practice Camp firings at towed 'sleeves', using kine-theodolites to measure the relative position of shell bursts and target. This was much too slow a target and the Practice Camp computational analysis was not very accurate. Later, in April 1940, a special trial of predictor accuracy was staged, following a free flying aircraft, and using camera recordings of the predictor output dials synchronised with kine-theodolites tracking the target. However, when German aircraft began to come over England later in 1940, it was at once clear that the aiming errors under operational conditions were much greater than those estimated from trials. We were up against the problem of increased operator inaccuracy under stress.

I remember P. M. S. Blackett (who was then in charge of the newly formed A. A. Command, Operational Research Group), wondering after watching the shell bursts in the night sky and a searchlight-held enemy aircraft, whether it would be possible to determine roughly an operational error distribution with appropriate photo-positioning equipment. I think that we later gave up all hope of estimating the actual aiming errors under operational conditions and made our calculations for a variety of different error combinations, which was often all that was needed in reaching conclusions about the relative

merits of different types of shell, etc. It was only towards the end of the war when we were faced with that ideal straight-line-flying target, the V1 flying bomb, and when using proximity fuzes that a rough operational check on the overall adequacy of the model could be made.

<u>THE FRAGMENTATION PROBLEM</u>. Before the war, the standard trials for determining the fragmentation characteristics of shell were:

(a) Fragmentation in a sand-bag 'beehive', the shell fragments being recovered, passed successively through various sizes of sieve and (above a certain minimum size) counted and weighed.

(b) Trials to measure the dispersion and penetrating power of fragments by detonating the shell some 5 ft. above ground, in a surround of 2-inch-thick wooden targets, placed in a semicircle of, say, 30, 60, 90 or 120 ft. radius. The detonation was either at rest or obtained by firing the shell with appropriate remaining velocities against a light bursting screen.

With the war-time allocation of additional scientific effort onto weapon lethality problems, the number of questions which were posed for answering was greatly increased. The shell and bomb fragment attack on many targets besides aircraft had to be considered. On the one side it was necessary to have means of projecting individual fragments of various sizes at known velocities, against a variety of targets. On the other it was important to know more about the size-velocity-directional pattern as well as the retardation of the fragments projected by a complete shell burst in flight.

As soon as forward planning is attempted it becomes necessary to generalise the characteristics of a weapon; in the case of A. A. shell the ultimate objective was to be able to predict the characteristics of the fragment distribution from

(i) the drawing board design,

(ii) a knowledge of the particular explosive filling to be used, and

(iii) for any desired forward velocity of the shell.

It became clear that the old form of trials mentioned in the first paragraph of this section was inadequate. When shells were burst in flight in a wood target surround the resulting pattern of perforations could not be accurately related to the pattern from a static burst, merely by adding the component forward velocity of the shell. Nor was it easy to link the distribution of fragment sizes from the sand-bag collection with the number of perforations in the wood, using any simple assumptions about velocities and retardations. The essential need was for more basic physical experimentation; without this we could not generalise.

Here we were lucky in getting help from a very skilled scientific team at our Safety in Mines Research Establishment at Buxton, who initiated a programme of research which gradually succeeded in disentangling the picture. Shells on which small letters were engraved in successive rings round the circumference were fired at rest, within a surround of strawboard, against which a large number of small velocity measuring screens were placed. Fragments subsequently collected and weighed could be identified with a particular zone of the shell, and velocities estimated either by direct measurement or more crudely from depth of penetration into the strawboard.

It then became clear that the initial velocity of fragments varied very considerably with the part of the casing from which they came and similarly, that size or weight also varied with position. To some extent this initial velocity could be related to the charge/weight ratio of the section of the shell (perpendicular to its axis) from which the fragments originated. With this information, we began at last to get a surer picture of how fragments would be projected from different designs of shell detonated at any given velocity in free air.

It should be noted that the angle of the fragment zone, in particular the rather sharply defined 'cut-off angle' or semi-vertical angle θ of the backward bounding cone of my Figure 2 became particularly important with the introduction of proximity fuzes. If the pattern of fuze functioning was not co-ordinated with that of fragmentation the shell might generally burst in positions relative to the target such that fragments were bound to miss the more vulnerable parts of the aircraft.

AIRCRAFT VULNERABILITY. In the earliest trials carried out shortly before the war, an aircraft and an arc of large 2-inch thick vertical wooden screens were placed beyond and on opposite sides of a small burster screen at which the shell (with percussion fuze) was fired at a prescribed velocity.

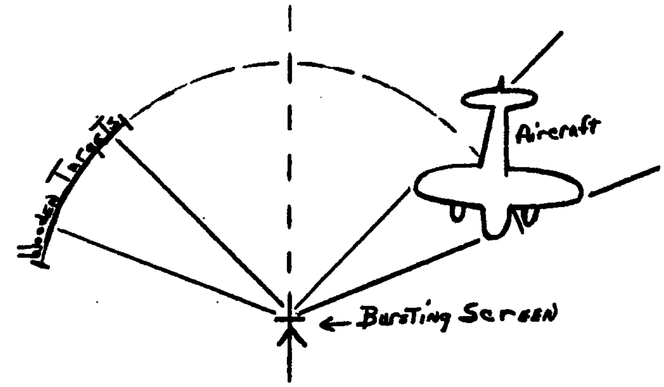


Figure 3

It was in this way possible to correlate the damage done to the aircraft with the density of fragments which perforated two inches of wood in a second, similarly constituted fragment stream. By noting and painting round the fragment holes after each round was fired, the same target could be used a large number of times, varying the aspect of attack and distance of detonation as desired.

It was from the observed correlation of density of 'throughs' (fragment capable of perforating 2 ins. of wood) and damage that it was possible to introduce into the model calculations a simplified 'equivalent vulnerable target'. This was the first method of attack. At a later stage after experimental techniques had become more refined and the Royal Aircraft Establishment assessors more experienced, it became possible to dissect the problem still further. The overall vulnerability picture was then built up from information gained by firing from high velocity barrels individual fragments of predetermined sizes, housed in specially designed cups, at a variety of aircraft components, which were screened where necessary by aluminium plates representing wing surfaces or fuselage.

The information so obtained could of course be used directly both in trying to draw conclusions about optimum fragment sizes and velocities and in considering ways of improving the protection of our own aircraft. Viewed in this way the problem may not appear to be statistical at all, but it did assume a statistical character as soon as one had to try and make use of this information in the 'model', with its shells bursting in a probability distribution around an aircraft and each projecting a composite stream of fragments, whose frequency distribution of strikes on equal areas of an intervening target would be roughly of Poisson form.

SOME CONCLUSIONS DRAWN FROM THIS SURVEY. Looking back now after a number of years, it seems to me that by 1944 we had really broken the back of the problem. It became possible to make recommendations with some confidence of a number of matters: on the optimum design characteristics of time fuzed and of proximity fuzed shell; on the relative importance of case thickness and explosive filling; on what might be achieved by using methods to control the size of fragments; on the relative gains to be won by improvement in fire control and in design of shell. Few such questions could have been answered with any confidence in 1939.

It is of course a truism that much of the fundamental research bearing on military problems is only rounded off when it is becoming too late to be of use in the war which provided the stimulus for the effort; and by the next war, the whole conditions of warfare are changed. This seems particularly true in regard to the ground-to-air weapons. But I think that the work I have been describing brought to the front a number of general principles, a sample of which I will bring to your attention in concluding this account.

The ease with which important factors may be overlooked. A common experience when the human mind starts to investigate the unknown is the way in which important considerations which seem so obvious afterwards are only realised through a process of slow and perhaps painful discovery.

(a) We did not for long appreciate the effect of ground ricochet in our firing trials. The influence of ricochet and other factors arising from proximity to the ground on the directional distribution of fragments would be natural operational effects in the case of field artillery or dropped bombs, but were very confusing when we were seeking information about the character of shell-bursts thousands of feet above ground. I know that the American experimenters appreciated this effect before we did and were the first to introduce ricochet traps into A. A. shell trials. Perhaps the most convincing demonstration of its existence which I recall occurred when we burst a 500 lb. bomb statically, with axis inclined at 30° to the vertical. The target screens showed a striking pattern of holes; a tilted belt like the forward-arm of a V from direct hits and another, like the other arm, from the ground ricochets. As long as bombs or shell were burst with their axes horizontal (or vertical), the effect remained unnoticed.

(b) Again, when studying the size distribution of fragments, the amount of secondary break-up on striking the collecting medium after detonation, was only realised when strawboard was used in place of sand

and the paths of these pieces, broken on first strike, could be traced through the successive layers of board.

(c) Another point not fully appreciated was the effect of emotional stress on the human element under battle conditions. The assessment of its magnitude, especially under circumstances and conditions which cannot be precisely foretold, is one of the hardest problems of the moment.

The place of basic research. In many instances it may not be too difficult to carry out a realistic trial of a particular weapon, against a given target under specified environmental conditions. But a more fundamental knowledge is necessary to assess the performance of weapons, perhaps still on the drawing board, under a wide variety of conditions. It was in this connection that the detailed experimental work on fragmentation performed to laboratory standards was essential, even if the laws of initial velocity, of size distribution and of retardation which resulted were to some extent empirical.

The value of having something up your sleeve. Observation of the amount of the metal casing which appeared to be broken up into dust or very small fragments*, on detonation, suggested that the destructive power of the anti-aircraft shell might be considerably increased by 'controlling' the size of fragments. It was over this matter that the help of the Safety in Mines Research Establishment was first called on, and by the end of the war this research group had developed a variety of techniques, relatively easily applied, by which it was possible to control the size and shape of shell and bomb fragments to a remarkable degree. These techniques were never used** but they were available to put into operation should any new target have had to be faced, e.g. a tough one against which only large fragments could be effective.

* It was realised later that some of this effect was due to secondary break-up of the large fragments on striking the collecting medium.

** It was found later that the Germans had applied a system of external grooving to some of their A. A. shell, apparently to increase the fragment size.

These are some of the still relevant points which I have noted in again coming into contact with problems of weapon research and development after a gap of several years. I am sure there are other lessons to be drawn from these World War II investigations, and without doubt those scientists who have carried on continuously in government service will have quietly absorbed them, so that they form part of their whole attitude of approach to the problems of today.

THE POSITION TODAY. There are, of course, many obvious differences between:

(a) The war-time problem, which was essentially that of trying to establish an understanding of a weapon system in service, in order to determine how its effectiveness could be improved, under conditions which were not expected to change radically from those known to exist; and

(b) The problem of today, which is greatly concerned with predictive assessments of the operational performance of future systems, taking many years to develop and to be used against an opponent whose future equipment, weapons and tactics must be to a large extent a matter of guesswork.

In the course of war, even when action has to be taken to meet a new situation, this can be done by working on the basis of information which possesses some element of reality. A good example of this occurred in 1944 with the launching of the V.1 flying bombs against London. Within a few days a complete bomb which had been shot down without exploding was recovered, and immediate steps could be taken to estimate its vulnerability to shell-fire and fighter attack.

As far as I can recall, priority trials were undertaken to determine (a) the burst pattern of a proximity fuze around such a target, and (b) the nature and extent of its vulnerability to A. A. shell fragments. How quickly we went as far as inserting these new parameters into our probability model, I cannot remember; but it must have been soon evident that the V.1 was a target which could be successfully engaged by 3.7 inch anti-aircraft guns with existing shell, provided they were supplied with proximity fuzes. The large-scale delivery of American fuzes and the appropriate re-deployment of guns, when achieved after some weeks when the fighter aircraft had been forced to take the leading defence roll, played a very large part in countering the menace.

The scientific effort, when it became accepted as of value by the armed services, was quite naturally first directed to the study of the performance of individual weapons or pieces of equipment; the radar set, the proximity fuze, the terminal ballistics of a shell or of a variety of anti-tank weapons. Today there is a special demand for scientific aid in the intractable job of peering into the future. The lead for this activity was of course provided by the Operational Research Sections which were closely associated with various operational commands during the war. In this very difficult field of prediction in which the last war's operational experience becomes less and less relevant, the scientific line of attack must consist in welding together a great number of elements.

The following scheme of relationships illustrates what I mean by the many-sided approach:

Mathematical models Paper study	$\rightleftarrows$	Laboratory experiments Simulator trials Range trials
War games Computer games	$\rightleftarrows$	Field experiments Army exercises

The overall inferences to be drawn from the whole build-up are not of course matters of statistics; but the use of the theory of probability and of stochastic processes is implicit in the studies on the left-hand column, while statistical planning plays its part in the laboratory experiments and the range trials--even to some extend in the field trials.

I have already tried to illustrate the great value of a mathematical model in forming the structure against which an evaluation problem may be broken up into parts for separate study. In so doing attention is drawn to the links in the construction where essential information in quantitative form is most needed and perhaps most lacking. Again, and this is important, by permitting a good deal of elasticity in the mechanism and allowing for the introduction of factors which might conceivably operate in a future situation, the model may be used to extrapolate beyond the envelope of engagement conditions tested during field trials or even accepted as likely under present combat conditions.

The application of the model approach to the problem of ground-to-air missile evaluation is the natural successor to the war-time investigations which I have described. The break-up of the problem for study under four headings still remains as before.

(a) Engagement geometry, (b) fuze performance,

(c) warhead effectiveness, (d) target vulnerability.

But problem (a) has taken a much more complex shape, involving perhaps the use of both analogue and digital computers. The war game has an essential part to play as a research tool in the combined attack on the problem of developing weapons, equipment and tactics for the future. Its main function is perhaps to aid thought and analysis rather than to obtain directo results. By injecting the human decision process into the study, it provides an insight into the complex nature of land battle which it would be hard to get in any other way. In this form of study, as elsewhere, the essential need to formulate rules, focuses attention on the limiting conditions which have to be accepted by whatever route we try to make predictions of the performance of future systems.

As a final illustration of where we now stand, let me refer to a problem of considerable present interest in whose solution a number of the techniques tabled above might be called in. This is the problem of comparing the merits of the free flight gun and the guided missile in the ground attack on armour. Both types of weapon depend, though in very different degree, on the human operator:

The free flight gun. Here we have a system, fairly well understood which has been studied for years and for which a reasonable idea of performance under operational conditions is available. The operator has only to concentrate while laying the gun and, after firing, plays no further part in the fate of that particular round. The greatest element of uncertainty lies in the vulnerability of his own gun, and to assess this requires rather extensive study of visibility and audibility in a variety of environments.

The guided weapon. The advantage of this weapon is that its firing position can be concealed behind the crest of a hill. However, the human controller who must see the target, has to concentrate for a considerable time (depending on the range) in guiding the weapon onto the target. That he can do this with fair success has been demonstrated on a simulator and with live weapons used under trial conditions. The open question here is whether he can maintain this performance in an operational setting, when subject to the fears and emotions to which he would be exposed in battle.

A sound basis for any policy decision on these alternative systems must depend on a comparative quantitative assessment; this cannot be completed

without these missing pieces of information--the vulnerability of the gun to enemy counter action, the fall-off in human performance in a battle setting, and now adding to the puzzle, the observational power of the helicopter. Success in solution depends not only in not overlooking these considerations but, in persuading authority to provide the means of proceeding to the answers. How often one wonders have important decisions on weapon policy had to be taken in the past when the basic information for a real comparison was not available, although with greater foresight, perhaps, it might have been obtained in time.

Finally, it may again be asked: what of the statistician? Have I pushed him out of the picture: I think not. You must remember that I have been concentrating on a particular aspect of this matter of research, development and testing--the assessment of operational performance of weapons. In this peculairly difficult field, the statistician becomes the scientist who must merge his statistical identity into that of a group of men trained in several disciplines, but prepared to give no undue weight to any one of them in searching for answers to the problems in hand. That at any rate has been my personal experience.

APPLICATION OF 2^{8-4} FRACTIONAL FACTORIALS IN SCREENING OF VARIABLES AFFECTING THE PERFORMANCE OF DRY PROCESS ZINC BATTERY ELECTRODES

Nicholas T. Wilburn
U. S. Army Electronics Research and Development Laboratory
Fort Monmouth, New Jersey

Among its research and development activities on zinc-silver oxide batteries for special applications, USAELRDL is investigating the preparation of zinc electrodes by dry processes. These involve the application of dendritic zinc powders under pressure to the grids. The interlocking properties of the dendritic zinc particles make it possible to form the electrodes with moderate pressures such that the porosity and related high surface area of the electrode is not destroyed. It is expected that dry process zinc electrodes will have many advantages over conventional electrodeposited sponge zinc electrodes, including higher discharge efficiency, greater uniformity of performance and better adaptability to mechanized production with resultant economics.

Due to the large number of variables affecting the discharge performance of the electrodes, it was decided to design and conduct a fractional factorial experiment to isolate the significant variables. These variables, and any controlling interactions between them, could then be studied further to arrive at the optimum conditions for the production of electrodes of maximum discharge efficiency. The fractional factorial experiment was thus intended for the preliminary screening of all major variables acting simultaneously. As such, it was recognized that it is the most efficient and economical process known for accomplishing this, in addition to providing valuable data on interactions between the variables which cannot be obtained from the more widely used one or two at a time variable investigations.

There were two categories of variables in the electrode investigation, those related to the electrolytic formation of the dendritic zinc powders and those related to the electrode preparation itself. Although several more variables were considered, it was decided to limit the number of variables to eight, keeping all other factors constant. The eight selected variables are shown in Figure 1. High and low levels as shown were assigned to each variable. Although considerable thought was given to determination of the levels, it is seen in retrospect that wider ranges might have been assigned in some instances. Based on available literature, these were, however, considered sufficiently wide ranges.

---

*Figures can be found at the end of this article.

Variable G, the electrolyte temperature during the plating operation, turned out to be impossible to control with the plating equipment which was prepared for the experiment and with the selected plating current densities. The experiment thus became one involving seven variables, each at two levels. Rather than to eliminate G, however, it was felt that this paper would serve a broader purpose if it gave the fractional factorial procedures for the study of eight as well as seven variables. Variable G should, therefore, be considered both in and out of the figures and analysis.

Having established the variables and the high and low levels, the fractional replicate design was established, as shown in Figure 2. This is the first design, a second having been run later for reasons to be discussed. The design involves eight variables (or seven) each at two levels to be studied with a total of sixteen different electrodes. A full 2^8 factorial experiment, eight variables each at two levels, would involve 2^8 or 256 trials (128 trials for a 2^7 factorial). Therefore, this design represents a one sixteenth fraction of the 2^8 factorial (one eighth of the 2^7), expressed as a 2^{8-4} design (or 27-3 for seven factors). The design is based on extension of a basic 2^4 or 16 trials. Thus the first four variables are arranged in standard order for the 2^4 factorial. The other variables are then introduced by making a basic assumption that three factor interactions between the first four variables are negligible. Therefore E is introduced by equating it to the interaction between variables A, B, and C. Regarding the high level as plus and the low level as minus, the level of E for the first box is $- \times - \times - = -$, $+ \times - \times - = +$ for the second box, etc. Similarly, variable F is introduced to equating it to the BCD interaction, variable G to the ABD interaction and variable H to the ACD interaction. Normally in a seven variable design, the seventh variable would be equated to ABD instead of ACD as was done here since variable G was dropped out. However, this does not affect the 2^{7-3} experiment in any way.

Having thus established the fractional design, the sixteen electrodes were prepared in accordance with the high and low level criteria for each variable. The electrodes were then tested one at a time under standard and carefully controlled conditions to give the sixteen yields or responses as shown. The response is the discharge efficiency of the electrode expressed in percent as the ratio of the output capacity (at a discharge current density of 1.1 amp/square inch to a 0.3 volt change for the electrode) to the theoretical capacity of the zinc active material. For comparison the conventional sponge zinc electrodes give an average efficiency

of about 20% under comparable discharge conditions. The average of the sixteen responses is 31.2%. Before proceeding with the analysis of the responses, it should be noted how the design can be used to relate the sixteen mean effects obtained from the response analysis to the variables and their two factor interactions. Consider the high level boxes in the first four columns. The first mean effect, no high levels, will be for twice the average. The second will be for variable A, the third for B, the fourth for the AB interaction, the fifth for C and so on. The eighth (A, B and C at the high level) will be for variable E which was originally equated to ABC. Similarly the twelfth for G, the fourteenth for H and the fifteenth for F. Since this is a one sixteenth fraction, each of these principal effects will be confused with fifteen other effects. However, these for each of the eight variables will be three factor or higher order interactions which are considered negligible, the basis on which the design was established. Each two factor interaction, AB for example, will be confused with fifteen other effects of which three are other two factor interactions. The sixteenth row of the First Design will represent such a combination of four 2 factor interactions.

The analysis of the sixteen responses is shown in Figure 3. The technique used here is the Yates' Algorithm which is a rapid method for obtaining the same mean effects that would be obtained from a formal and lengthy analysis of variance. The Yates' Algorithm is applicable to any factorial experiment. Its advantages become more apparent the larger the experiment.

The mechanics of the Yates' calculations are very simple. The first figure in Column (1) is the sum of responses 1 and 2; the second, the sum of responses 3 and 4, etc. The ninth figure is the sum of responses 1 and 2 with the sign of response 1 reversed. Column (2) is derived from Column (1) in the same manner. Additional columns are introduced until a column is completed the first figure of which is equal to the sum of the responses. The arithmetic in each column is checked before proceeding to the next column. The sum of Column (1) is equal to twice the sum of the even numbered responses; the sum of Column (2) is equal to four times the sum of every second even numbered response; the sum of Column (3) is equal to eight times the sum of every fourth even numbered response; and the sum of Column (4) is equal to sixteen times the sixteenth response. The mean effects are obtained by dividing each figure of Column (4) by eight. The sum of the mean effects is also checked. The 62.363 figure, twice the average, is not used in the subsequent analysis. The

effects measured by the mean effects are given in the last column. The effects A, B, AB, C, AC, etc. are those as previously read off the first design chart. The three factor and higher order interactions which they are confused with are of no importance since the design is based on their being assumed negligible. The other two factor interactions are important and must be known. They are obtained from the effects A, B, AB, etc. and the defining contrasts of the design. These defining contrasts are obtained from the equating that had been done in establishing the fractional design; E = ABC, F = BCD, G = ABD and H = ACD. Sixteen defining contrasts are required for the eight variable design. The first defining contrast is always I, the next four are ABCE, BCDF, ABDG and ACDH. The remaining eleven are found by exhaustively multiplying these contrasts using the rule that like factors are cancelled out.

(1) I
(2) ABC x E = ABCE
(3) BCD x F = BCDF
(4) ABD x G = ABDG
(5) ACD x H = ACDH
(6) $A\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}DF = ADEF$
(7) $\cancel{A}\cancel{B}CE \times \cancel{A}\cancel{B}DG = CDEG$
(8) $\cancel{A}B\cancel{C}E \times \cancel{A}\cancel{C}DH = BDEH$
(9) $\cancel{B}C\cancel{D}F \times A\cancel{B}\cancel{D}G = ACFG$
(10) $B\cancel{C}\cancel{D}F \times A\cancel{C}\cancel{D}H = ABFH$
(11) $\cancel{A}B\cancel{D}G \times \cancel{A}C\cancel{D}H = BCGH$
(12) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}B\cancel{D}G = BEFG$
(13) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}C\cancel{D}H = CEFH$
(14) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{A}\cancel{B}\cancel{D}G \times A\cancel{C}\cancel{D}H = AEGH$
(15) $\cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}\cancel{B}\cancel{D}G \times \cancel{A}\cancel{C}DH = DFGH$
(16) $\cancel{A}\cancel{B}\cancel{C}E \times \cancel{B}\cancel{C}\cancel{D}F \times \cancel{A}B\cancel{D}G \times ACDH = ABCDEFGH$

(These sixteen defining contrasts may also be obtained by simply reading off the low level boxes for each of the sixteen trials in the First Design chart. However this procedure will not apply in all cases, e.g. in the 2^{7-3} it would give sixteen defining contrasts when only eight are required.)

The principal effects are found by multiplying the first effect, A, B, AB, etc., by the sixteen defining contrasts. The two factor interactions confused with AB are found for example to be CE, DG and FH.

$$AB \times I = AB$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}CE = CE$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}DG = DG$$
$$\cancel{A}\cancel{B} \times \cancel{A}\cancel{B}FH = FH$$

The other twelve effects confused with AB, all three factor or higher order interactions, could be found with the other twelve defining contrasts if desired. The total 256 effects found by multiplying each starting effect by the sixteen defining contrasts would give the full 256 effects, from I to the interaction ABCDEFGH, which would be obtained in conducting a full 2^8 factorial experiment.

The defining contrasts for the 2^{7-3} experiment are I, ABCE, BCDF, ACDH, AEDF, BDEH, ABFH, and CEFH. The effects measured are found in the same way as for the eight variable design. The only difference is that, in this case, all effects involving G drop out. The twelfth set of effects measured, identified by the asterisk, then becomes eight interactions, all three factor or higher order.

The fifteen mean effects with their identifying effects are then ordered by arranging them in order of magnitude without regard to sign. The thus ordered set of mean effects is then arranged in a half-normal plot as shown in Figure 4 to interpret the relative significance of the effects. The ordinate is the order number of the fifteen effects from smallest to largest. The abscissa gives the mean effect magnitudes. The fifteen points which are plotted are identified by the proper major effects and two factor interactions. The plot is given for both the full eight variables, and the actual seven variable experiment. In the latter case, the G factor and the G interactions drop out. The asterisk again denotes high order interactions. In the plot an error best straight line has been drawn through the lowest seven points. High magnitude effects falling significantly off the line are judged to be distinct from error and therefore controlling factors in the process being investigated. In a plot of this type, it is considered unusual, however, to have a well defined error line with as many as eight points falling clearly off it. To gain insight into this unusual behavior as well as to gain more precision in the estimation of all of the effects, it was decided to conduct a second phase of the experiment, another group of sixteen electrodes differing from the first. Since an interaction between A (the highest magnitude effect) and B appeared reasonable, and since F and H were both high magnitude effects which might interact with each other, it was decided to establish the second

design in such a way as to separate the AB and FH interactions. Although this can be done in several equally effective ways, it was decided to do it by reversing the levels for factors A and D, giving the design as shown in Figure 5.

This design is identical to the first except for the reversal of levels of variables A and D. Sixteen electrodes were prepared in accordance with the indicated variable levels. The electrodes were then tested to give the listed responses. The next step in the experiment was to determine the mean effects generated by these responses and to combine them with the mean effects resulting from the first design. This may be done in two ways. The first is to combine the two sets of sixteen responses into an overall group of thirty-two and then to conduct a Yates' computation to arrive at the thirty-two mean effects. The second way is to conduct a Yates' computation on the second design responses and then combine the mean effects with those of the first design. The first method is less time consuming and therefore preferable. The second method gives a clearer picture of the separation of effects and is therefore now given. (The first method is given in Appendix A-1)

Figure 6 shows the Yates' computations on the responses from the second design. The operations are identical to those as described for the first design. It is noted that the signs of all elements involving A and D in the Effects Measured column are now minus since the levels of A and D had been reversed. This reversal of signs permits the separation of effects as shown in Figure 7.

The mean effects derived from the first design are given in the column marked X. Those derived from the second design are given in the column marked Y. An example of the computation to separate the effects is as follows:

The second mean effect in the X column is for variable A. The similar mean effect in the Y column is for minus variable A with the difference in the absolute magnitudes of the mean effects being due to experimental error. Reversing the sign of the column Y mean effect and averaging it with that of the column X mean effect will give the -4.81 mean effect for variable A and certain high order interactions confused with it. 1/2(X+Y), -1.36, then gives the mean effect for the remaining high order interactions originally confused with A. Similar calculations are performed to separate all of the other effects, those containing A or D from each of the others. The thirty-one statistics

thus obtained, not including the 61.22 (twice the average) figure, are then arranged in order of magnitude without respect to sign and plotted in a thirty-one factor half-normal plot as shown in Figure 8.

It is seen that the error best straight line is now established by twenty-three of the thirty-one points reflecting the greater precision achieved by doubling the experiment and thereby reducing the variance of the estimated effects by one-half. Of the eight points which are clearly off the line, two of them, denoted by asterisks, are combinations of high order interactions. Their relative significance cannot be interpreted within the limits of the experiment. As will be seen, equating them to zero will not affect the results. The BF interaction is not far off the line and its significance may be questioned. Variables E and A are both clearly controlling factors in the efficiency of the zinc electrodes. The signs of their mean effects are both minus, indicating that higher efficiencies can be obtained at the lower levels of the ranges studied, in other words at the lower pressing temperature, 80° F, and with the smaller weight of zinc per plate. The interpretation of variables H, the formation current density, and F, the presence or absence of zinc oxide in the formation electrolyte, is more complex. This results from the probable significance of FH, the interaction between them (CE is very unlikely to be significant due to the low magnitude of C). In general when an interaction is large, as in this case, the corresponding mean effects cease to have much meaning. The effect of F is clearly dependent upon the level of H and vice versa. The three effects F, H and their interaction FH may best be interpreted as a single highly significant effect. Further experimental work at intermediate levels for the two variables is definitely indicated.

The final step in the analysis was to determine if the conclusion was correct that only the effects E, A and the combined effects of H, F and FH were significant, i.e., distinct from experimental error. Part of the purpose of this final step was to determine if the entire experiment was valid, in other words, that there were no large errors made in the actual responses which could have seriously altered the mean effects. The procedure used was to determine the standard error of the individual observed responses by analyzing the thirty-one mean effects. A second standard error, for the differences between observed and predicted responses, was then obtained with the predicted responses based on the assumption that all mean effects other than those for E, A, H, FH and F were indeed zero. If the two standard errors would then be equivalent, both the total experiment and the conclusions derived from it would be proved valid.

Since the error straight line on the final half-normal plot was established by the twenty-three lowest magnitude points, a half-normal plot for these points was prepared. The standard error for the individual observed responses as derived from this plot was 3.0. (See A-5). All mean effects other than the twice the average effect, E, A, H, FH and F were then equated to zero and a reverse Yates' computation was conducted to obtain the thirty-two predicated responses (See A-2). These responses were compared with the observed responses and a list of the thirty-two differences between the observed and predicted responses was prepared (See A-3). The magnitudes of the differences were plotted on normal probability paper. A standard error was obtained from this plot for the difference between individual observed responses and individual predicted responses. The value of this standard error was 3.2 (See A-4). This was in excellent agreement with the standard error of the individual observed responses, thus proving the validity of the experiment and the conclusions derived from it.

In conclusion, a fractional factorial designed experiment, involving two 2^{7-3} fractional designs, has been conducted to determine the significance of seven major variables, and two factor interactions between them, on the discharge efficiency of the dry process zinc battery electrodes. A total of thirty-two electrodes was prepared and tested. Analysis of their responses has indicated the controlling influence of two of the variables, pressing temperature and the amount of zinc per plate, and of the interaction between two other variables relating to the plating conditions under which the zinc material was prepared. The other three variables have been shown to be unimportant in comparison, within the range of levels selected. The experiment has fulfilled its basic purpose, narrowing down the range of variables to permit extensive investigation of the truly important variables in order to arrive at the optimum electrode preparation procedures in the most expeditious manner.

Figure 1

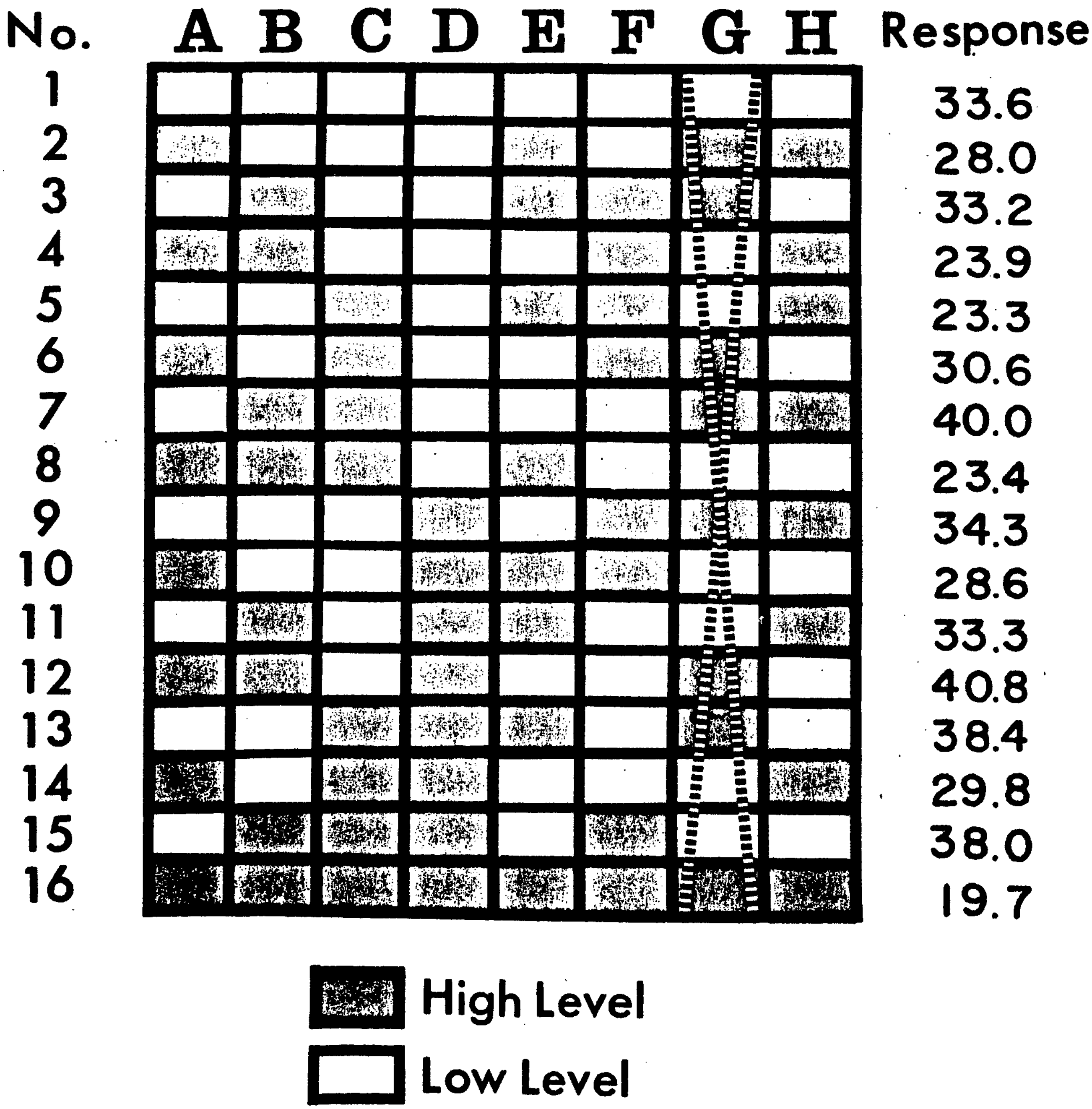
Experiment Variables

	Pressing Variables	Units	High	Low
A	Zinc weight	grams	5.23	2.62
B	Pressure	psi	1,840	1,230
C	Particle size	sieve mesh	100	200
D	Pressure time	minutes	15	1
E	Pressure temp.	°F	300	80
	Formation Variables			
F	ZnO in electrolyte	gm./liter	20	0
~~G~~	~~Electrolyte temp.~~	~~°F~~	~~100~~	~~80~~
H	Current density	amp./sq. in.	1.0	0.75

Figure 2

FIRST DESIGN

Variables



FIRST DESIGN ANALYSIS, YATES' ALGORITHM

NO.	RESPONSE	(1)	(2)	(3)	(4)	MEAN EFFECTS (4)/8	EFFECTS MEASURED
1	33.6	61.6	118.7	236.0	498.9	62.363	-
2	28.0	57.1	117.3	262.9	-49.3	-6.163	A
3	33.2	53.9	137.0	-24.2	5.7	0.713	B
4	23.9	63.4	125.9	-25.1	-24.1	-3.013	AB + CE + ~~DG~~ + FH
5	23.3	62.9	-14.9	5.0	-12.5	-1.563	C
6	30.6	74.1	-9.3	0.7	-23.1	-2.888	AC + BE + DH + ~~FG~~
7	40.0	68.2	1.8	-27.6	-7.7	-0.963	AE + BC + DF + ~~CG~~
8	23.4	57.7	-26.9	3.5	-43.1	-5.388	E
9	34.3	-5.6	-4.5	-1.4	26.9	3.363	D
10	28.6	-9.3	9.5	-11.1	-0.9	-0.113	AD + ~~BG~~ + CH + EF
11	33.3	7.3	11.2	5.6	-4.3	-0.538	~~AG~~ + BD + CF + EH
12	40.8	-16.6	-10.5	-28.7	31.1	3.888	~~G~~*
13	38.4	-5.7	-3.7	14.0	-9.7	-1.213	AH + BF + CD + ~~EG~~
14	29.8	7.5	-23.9	-21.7	-34.3	-4.288	H
15	38.0	-8.6	13.2	-20.2	-35.7	-4.463	F
16	19.7	-18.3	-9.7	-22.9	-2.7	-0.338	AF + BH + ~~CG~~ + DE
CHECKS	498.9	449.6	431.2	344.8	315.2	39.396	

449.6
431.2
344.8
315.2
39.400

*3 factor and higher order interactions

Figure 3

Figure 4

Half-Normal Plot, First Design

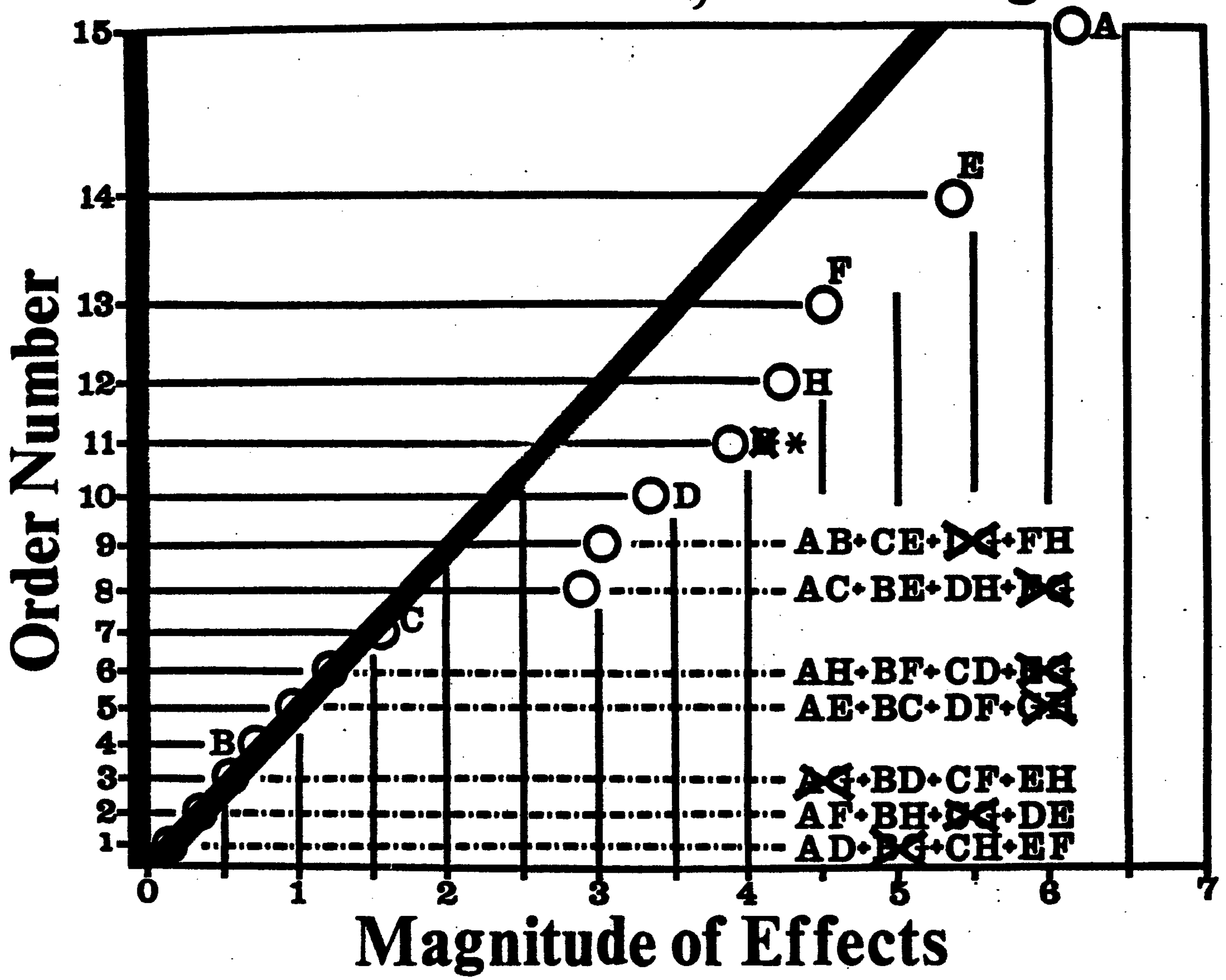
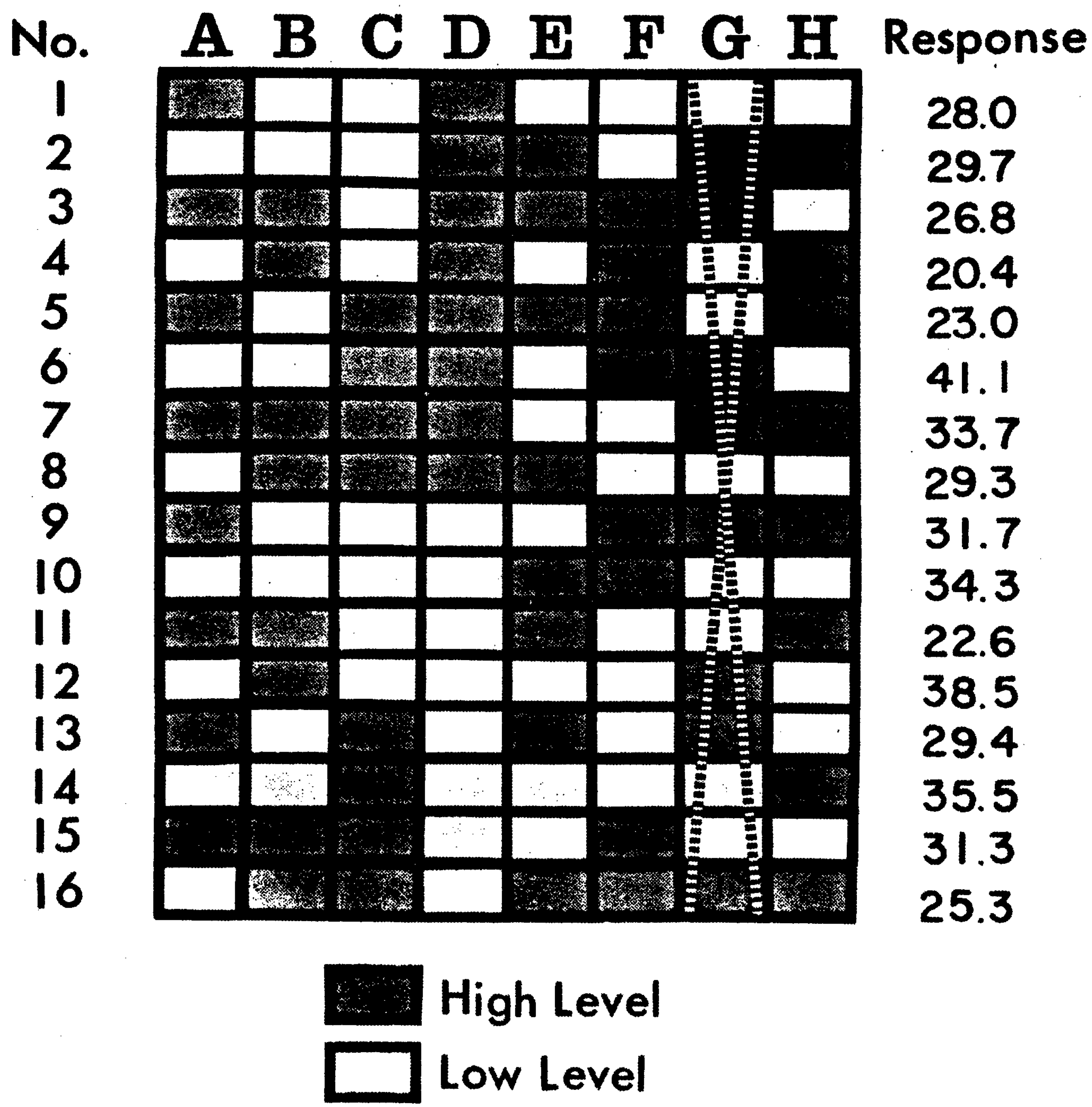


Figure 5

SECOND DESIGN

Variables



SECOND DESIGN ANALYSIS, YATES' ALGORITHM

NO.	RESPONSE	(1)	(2)	(3)	(4)	MEAN EFFECTS (4)/8	EFFECTS MEASURED
1	28.0	57.7	104.9	232.0	480.6	60.075	-
2	29.7	47.2	127.1	248.6	27.6	3.450	-A
3	26.8	64.1	127.1	9.0	-24.8	-3.100	B
4	20.4	63.0	121.5	18.6	-29.4	-3.675	-AB + CE - ~~DG~~ + FH
5	23.0	66.0	-4.7	-11.6	16.6	2.075	C
6	41.1	61.1	13.7	-13.2	0.0	0.000	-AC + BE - DH + ~~BG~~
7	33.7	64.9	18.5	-30.6	6.0	0.750	-AE + BC - DF + ~~CG~~
8	29.3	56.6	0.1	1.2	-39.8	-4.975	E
9	31.7	1.7	-10.5	22.2	16.6	2.075	-D
10	34.3	-6.4	-1.1	-5.6	9.6	1.200	AD + ~~BG~~ + CH + EF
11	22.6	18.1	-4.9	18.4	-1.6	-0.200	-~~AG~~ - BD + CF + EH
12	38.5	-4.4	-8.3	-18.4	31.8	3.975	~~G~~*
13	29.4	2.6	-8.1	9.4	-27.8	-3.475	-AH + BF - CD + ~~EG~~
14	35.5	15.9	-22.5	-3.4	-36.8	-4.600	H
15	31.3	6.1	13.3	-14.4	-12.8	-1.600	F
16	25.3	-6.0	-12.1	-25.4	-11.0	-1.375	-AF + BH + ~~CG~~ - DE
CHECKS	480.6	508.2	454.0	436.8	404.8	50.600	

508.2
454.0
436.8
404.8
50.600

*3 factor and higher order interactions

Figure 6

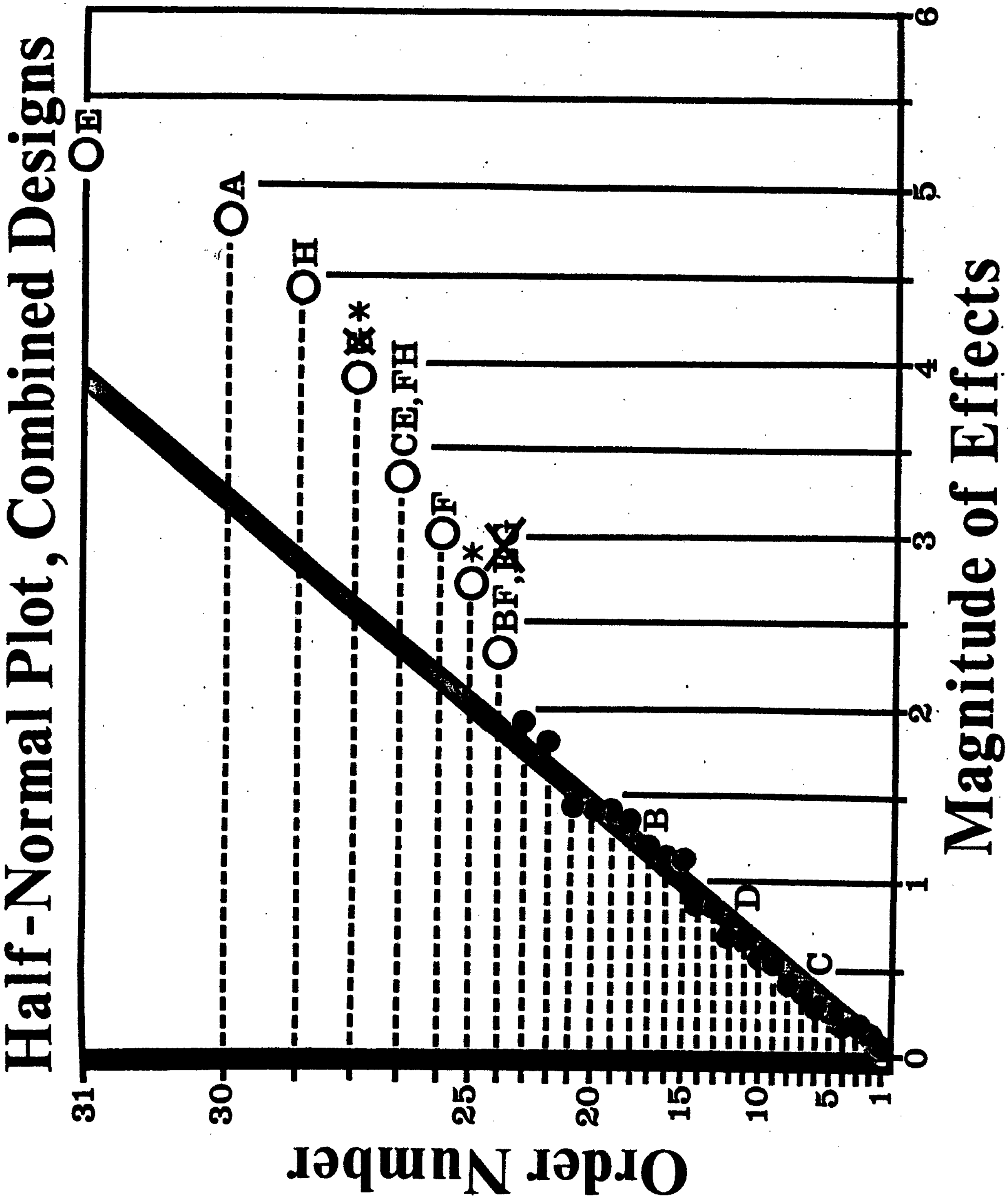
DETERMINATION OF EFFECTS, COMBINED DESIGNS

No.	X Mean Effects First Design	Y Mean Effects Second Design	½(X + Y)	Effects Measured	½(X - Y)	Effects Measured
1	62.363	60.075	61.22	-	1.14	Blocks
2	-6.163	3.450	-1.36	*	-4.81	A
3	0.713	-3.100	-1.19	B	1.91	*
4	-3.013	-3.675	-3.34	CE + FH	0.33	AB + ~~DG~~
5	-1.563	2.075	0.26	C	-1.82	*
6	-2.888	0.000	-1.44	BE + ~~EG~~	-1.44	AC + DH
7	-0.963	0.750	-0.11	BC + ~~GH~~	-0.86	AE + DF
8	-5.388	-4.975	-5.18	E	-0.21	*
9	3.363	2.075	2.72	*	0.64	D
10	-0.113	1.200	0.54	AD + ~~BG~~ + CH + EF	-0.66	*
11	-0.538	-0.200	-0.37	CF + EH	-0.17	~~AG~~ + BD
12	3.888	3.975	3.93	~~G~~*	-0.04	*
13	-1.213	-3.475	-2.34	BF + ~~FG~~	1.13	AH + CD
14	-4.288	-4.600	-4.44	H	0.16	*
15	-4.463	-1.600	-3.03	F	-1.43	*
16	-0.338	-1.375	-0.86	BH + ~~CG~~	0.52	AF + DE
			45.01		-5.61	
			-5.61			
			34.90			
			x 8			
			315.20	(Check)		

* 3 factor and higher order interactions

Figure 7

Figure 8


Half-Normal Plot, Combined Designs
Magnitude of Effects
Order Number
E
A
H
CE,FH
F
B
D
C
0
1
2
3
4
5
6
31
30
25
20
15
10
5
1

APPENDIX

This appendix contains supplemental data as follows:

A-1. Yates' Algorithm computation for thirty-two responses, combination of Designs 1 and 2.

A-2. Reverse Yates' Algorithm computation for thirty-two effects, assuming all effects are zero except those for average, E, A, H, FH and F.

A-3. Comparison of observed and predicted responses.

A-4. Probability plot to obtain standard error of individual differences between observed and predicted responses.

A-5. Half-normal plot of twenty-three mean effects to obtain standard error of individual observed responses.

A-1. Yates' Algorithm computation for thirty-two responses, combination of Designs 1 and 2.

No.	Responses	(1)	(2)	(3)	(4)	(5)	(5)/16	Effects Measured
1	33.6	61.6	118.7	236.0	498.9	979.5	61.22	-
2	28.0	57.1	117.3	262.9	480.6	-21.7	-1.36	*
3	33.2	53.9	137.0	232.0	-49.3	-19.1	-1.19	B
4	23.9	63.4	125.9	248.6	27.6	-53.5	-3.34	CE+FH
5	23.3	62.9	104.9	-24.2	5.7	4.1	0.26	C
6	30.6	74.1	127.1	-25.1	-24.8	-23.1	-1.44	BE+FG
7	40.0	68.2	127.1	9.0	-24.1	-1.7	-0.11	BC+GH
8	23.4	57.7	121.5	18.6	-29.4	-82.9	-5.18	E
9	34.3	57.7	-14.9	5.0	-12.5	43.5	2.72	*
10	28.6	47.2	-9.3	0.7	16.6	8.7	0.54	AD+BG+CH+EF
11	33.3	64.1	1.8	-11.6	-23.1	-5.9	-0.37	CF+EH
12	40.8	63.0	-26.9	-13.2	0	62.9	3.93	G
13	38.4	66.0	-4.7	-27.6	-7.7	-37.5	-2.34	BF+EG
14	29.8	61.1	13.7	3.5	6.0	-71.1	-4.44	H
15	38.0	64.9	18.5	-30.6	-43.1	-48.5	-3.03	F
16	19.7	56.5	0.1	1.2	-39.8	-13.7	-0.86	BH+CG
17	28.0	-5.6	-4.5	-1.4	26.9	-18.3	-1.14	-Blocks
18	29.7	-9.3	9.5	-11.1	16.6	76.9	4.81	-A
19	26.8	7.3	11.2	22.2	-0.9	-30.5	-1.91	-*
20	20.4	-16.6	-10.5	-5.6	9.6	-5.3	-0.33	-AB-DG
21	23.0	-5.7	-10.5	5.6	-4.3	29.1	1.82	-*
22	41.1	7.5	-1.1	-28.7	-1.6	23.1	1.44	-AC-DH
23	33.7	-8.6	-4.9	18.4	31.1	13.7	0.86	-AE-DF
24	29.3	-18.3	-8.3	-18.4	31.8	3.3	0.21	-*
25	31.7	1.7	-3.7	14.0	-9.7	-10.3	-0.64	-D
26	34.3	-6.4	-23.9	-21.7	-27.8	10.5	0.66	-*
27	22.6	18.1	13.2	-9.4	34.3	2.7	0.17	-AG-BD
28	38.5	-4.4	-9.7	-3.4	-36.8	0.7	0.04	-*
29	29.4	2.6	-8.1	-20.2	-35.7	-18.1	-1.13	-AH-CD
30	35.5	15.9	-22.5	-22.9	-12.8	-2.5	-0.16	-*
31	31.3	6.1	13.3	-14.4	-2.7	22.9	1.43	-*
32	25.3	-6.0	-12.1	-25.4	-11.0	-8.3	-0.52	-AF-DE
	979.5	957.8	885.2	781.6	720.0	809.6	50.62	

Checks 957.8
885.2
781.6
720.0
809.6
50.60

A-2. Reverse Yates' Algorithm computation for thirty-two effects assuming all effects are zero except those for average, E, A, H, FH and F.

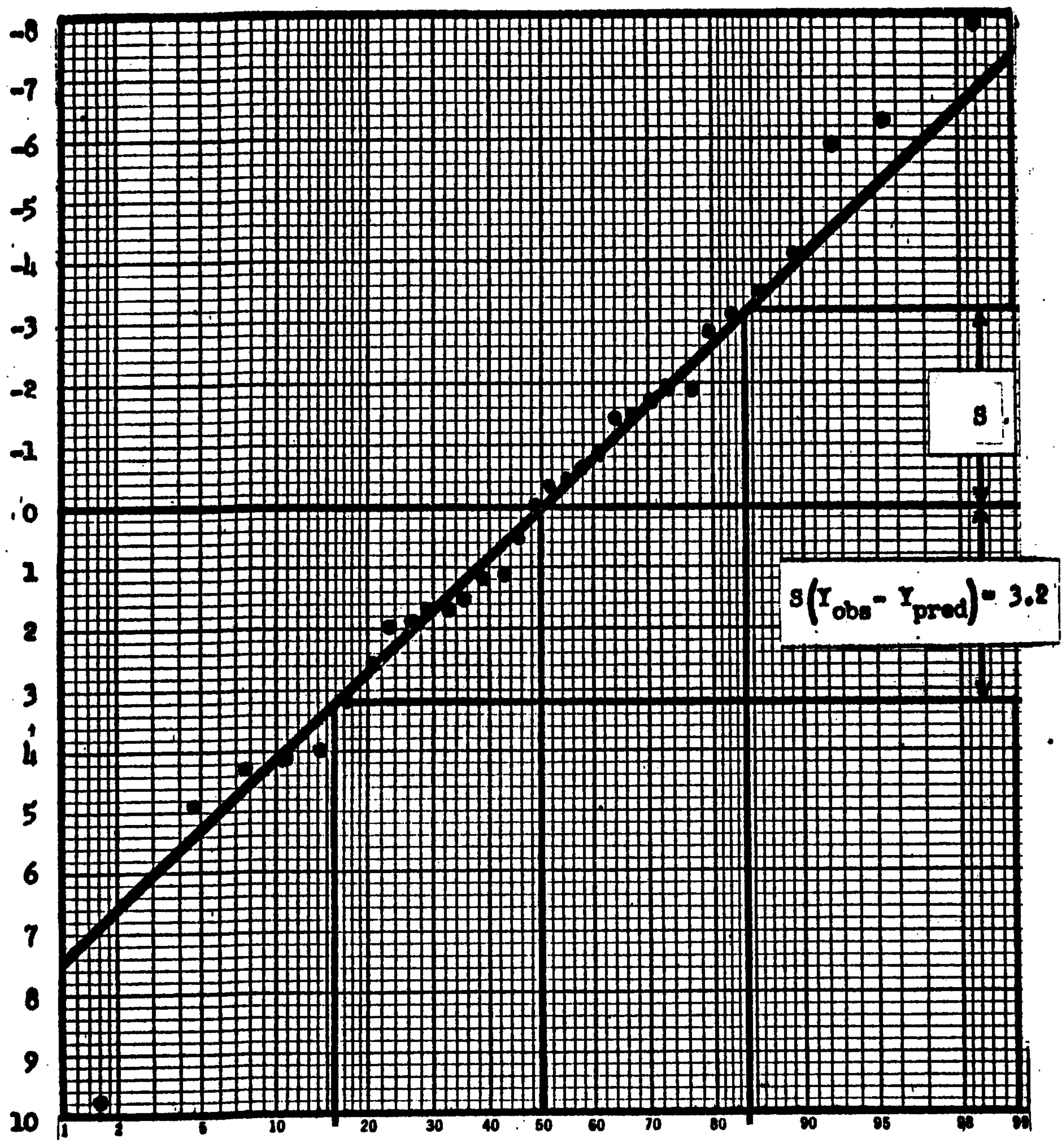
No.	Effects Measured	(5)	(5')	(6')	(7')	(8')	(9')	(10')	Predicted Response (10')/32
1	1	979.5	979.5	979.5	926.0	1,008.9	1,128.5	1,205.4	37.7
2	-	-21.7	0	53.5	-82.9	-119.6	-76.9	850.6	26.8
3	-	-19.1	0	0	0	-76.9	927.5	1,049.6	32.8
4	CE+FH	-53.5	-53.5	82.9	119.6	0	76.9	812.4	25.4
5	-	4.1	0	0	-76.9	950.1	972.7	800.4	25.0
6	-	-23.1	0	0	0	22.6	-76.9	1,061.6	33.2
7	-	-1.7	0	71.1	0	76.9	889.3	1,170.2	36.6
8	E	-82.9	-82.9	-48.5	0	0	76.9	885.8	27.7
9	-	43.5	0	-76.9	1,033.0	950.1	723.5	966.2	30.2
10	-	8.7	0	0	82.9	-22.6	-76.9	895.8	28.0
11	-	-5.9	0	0	0	-76.9	1,138.5	1,004.4	31.4
12	-	62.9	0	0	-22.6	0	76.9	1,051.6	32.9
13	-	-37.5	0	0	76.9	1,008.9	1,093.3	1,039.6	32.5
14	H	-71.1	-71.1	0	0	119.6	-76.9	1,016.4	31.8
15	F	-48.5	-48.5	0	0	76.9	962.7	1,215.4	38.0
16	-	-13.7	0	0	0	0	76.9	646.6	20.2
17	-	-18.3	0	979.5	1,033.0	843.1	889.3	1,051.6	32.9
18	-A	76.9	76.9	-53.5	82.9	119.6	-76.9	1,040.4	31.4
19	-	-30.5	0	0	0	-76.9	972.7	895.8	28.0
20	-	-5.3	0	-82.9	22.6	0	76.9	966.2	30.2
21	-	29.1	0	0	-76.9	1,115.9	972.5	646.6	20.2
22	-	23.1	0	0	0	-22.6	-76.9	1,215.4	38.0
23	-	13.7	0	-71.1	0	76.9	1,128.5	1,016.4	31.8
24	-	3.3	0	-48.5	0	0	76.9	1,039.6	32.5
25	-	-10.3	0	76.9	926.0	1,115.9	962.7	812.4	25.4
26	-	10.5	0	0	-82.9	22.6	-76.9	1,049.6	32.8
27	-	2.7	0	0	0	-76.9	1,093.3	850.6	26.6
28	-	0.7	0	0	-119.6	0	76.9	1,205.4	37.7
29	-	-18.1	0	0	76.9	843.1	1,138.5	885.8	27.7
30	-	-2.5	0	0	0	-119.6	-76.9	1,170.2	36.6
31	-	22.9	0	0	0	76.9	723.5	1,061.6	33.2
32	-	-8.3	0	0	0	0	76.9	800.4	25.0
			979.5 -48.5 931 x2 1,862	1,862 1,959 x2 3,918	3,918 x2 7,836	7,836 x2 15,672	15,672 x2 31,344	31,344	980.0 (979.5)

A-3. Comparison of observed and predicted responses.

No.	Observed Responses	Predicted Response	Obs. Resp.- Pred. Resp.	Ordered Series
1	33.6	37.7	-4.1	-9.8
2	28.0	26.6	1.4	-4.9
3	33.2	32.8	0.4	-4.3
4	23.9	25.4	-1.5	-4.1
5	23.3	25.0	-1.7	-4.0
6	30.6	33.2	-2.6	-3.2
7	40.0	36.6	3.4	-2.6
8	23.4	27.7	-4.3	-2.0
9	34.3	30.2	4.1	-1.9
10	28.6	28.0	0.6	-1.7
11	33.3	31.4	1.9	-1.7
12	40.8	32.9	7.9	-1.5
13	38.4	32.5	5.9	-1.2
14	29.8	31.8	-2.0	-1.1
15	38.0	38.0	0	-0.5
16	19.7	20.2	-0.5	0
17	28.0	32.9	-4.9	0.3
18	29.7	31.4	-1.7	0.4
19	26.8	28.0	-1.2	0.6
20	20.4	30.2	-9.8	0.8
21	23.0	20.2	2.8	1.4
22	41.1	38.0	3.1	1.5
23	33.7	31.8	1.9	1.7
24	29.3	32.5	-3.2	1.9
25	31.7	25.4	6.3	1.9
26	34.3	32.8	1.5	2.8
27	22.6	26.6	-4.0	3.1
28	38.5	37.7	0.8	3.4
29	29.4	27.7	1.7	4.1
30	35.5	36.6	-1.1	5.9
31	31.3	33.2	-1.9	6.3
32	25.3	25.0	0.3	7.9
	979.5	980.0	-44.5 +44.0 -0.5	

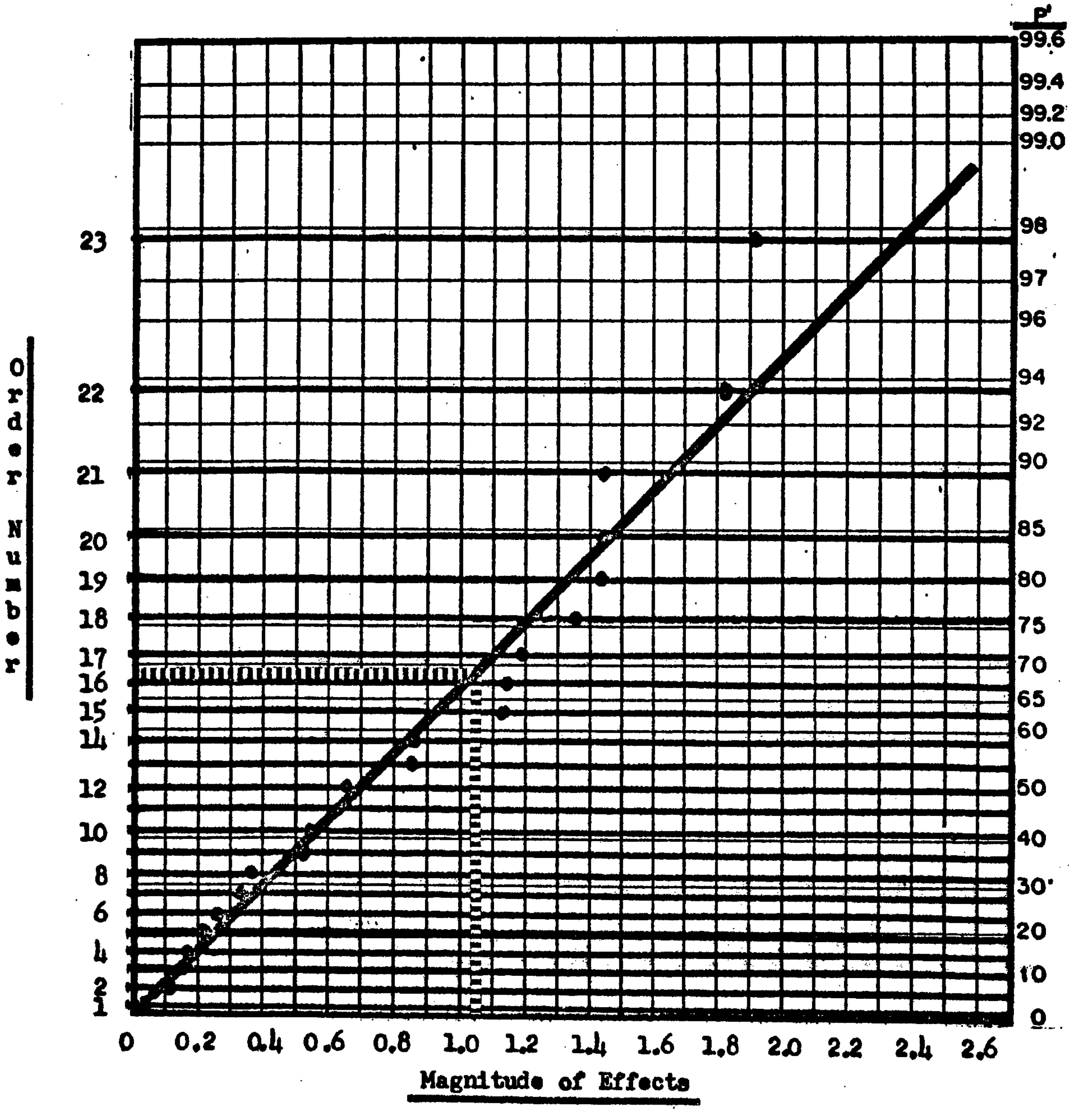
A - 4

Probability plot to obtain standard error of individual differences between observed and predicted responses



Probability Scale

A - 5 Half-normal plot of twenty-three mean effects to obtain standard error of individual observed responses.



$$S_Y = \frac{1.04}{\sqrt{\frac{1}{16}+\frac{1}{16}}} = 3.0$$

Applications of the Calculus of Factorial Arrangements
I. Block and Direct Product Designs

Badrig Kurkjian and Marvin Zelen
Harry Diamond Laboratories
and
Mathematics Research Center, U. S. Army, University of Wisconsin

ABSTRACT

This paper deals with some applications of a general theory for the analysis of factorial experiments as reported by the authors in the June 1962 issue of the Annals of Mathematical Statistics.

General expressions are given for the usual quantities associated with the analysis of variance for the cases where simple treatments or factorial treatment-combinations are applied to Randomized Blocks, Balanced Incomplete Blocks, Group Divisible designs, and a wide class of Kronecker Product designs.

The main point of the new theory is that, for a wide class of the more practical designs, the complete analysis can be carried out almost by inspection of the normal equations, with no requirement for inverting the normal equations.

The complete version of this paper is published in BIOMETRIKA, Vol. 50, Parts 1 and 2, June 1963.

STATISTICAL DESIGN OF EXPERIMENTS FOR CONTINUOUS DATA

Paul C. Cox
Reliability and Statistics Office
The Army Missile Test and Evaluation Directorate
White Sands Missile Range, New Mexico

I. INTRODUCTION. I wish to talk on the subject of how one would design an experiment and analyze the data when the results come in the form of a continuous curve, rather than just a single value. This is an area that would appear to have extensive application in science and engineering. For example: velocity data, trajectory data, meterological data, thrust data, etc. As I just mentioned, I am interested in the design of experiments, which means I am not concerned with the evaluation of a single curve, but many curves obtained as a result of testing under several sets of conditions, and very likely each set of conditions will have some replications.

To illustrate what I have in mind, I will use rocket motor thrust curves, although I could have used some other type of curve equally effectively. Now many results from rocket motor tests can easily be analyzed. For example: average exhaust velocity; effective (average) pressure; total impulse; specific impulse, etc. These are simple to analyze because the data for a given test usually comes in the form of one single number. However, if we want to estimate a typical or average thrust curve when a motor is tested under given conditions, this is quite a different problem.

To keep this report unclassified, the thrust data which will be discussed will be completely fictitious. The data is not, to my knowledge, appropriate for any existing rocket motor, but the general shape of the curve is similar to what may be expected for a number of motors currently in use.

The extensive information available in such areas as: Regression Analysis; Random Processes; Power Spectral Analysis; Time Series; Analysis of Covariance; Multivariate Analysis; and similar fields may easily cover the problem I am going to present. Therefore, my first question is, if the solution is readily available in the literature, I would (1)
like some references. My second question is, if it is not readily available but you know the answers, I will appreciate the information. (Please note (2)
that there are numbers along the margins of this report. These numbers refer to specific questions which may be found at the end of this report.)

I will conclude the introduction by stating that the five panel members, W. T. Federer, B. G. Greenberg, M. A. Schneiderman, H. L. Lucas, and H. O. Hartley have all prepared and forwarded their comments. These are included at the end of the report in the order in which they were received.

II. BASIC ASSUMPTIONS.

A. We will begin by assuming that the thrust curves to be considered in this study will look something like the one illustrated in Figure 1, although an actual curve will be somewhat more irregular than the example. The letters along the curve will represent the points which we will consider critical, and we will refer to these points many times in the future.

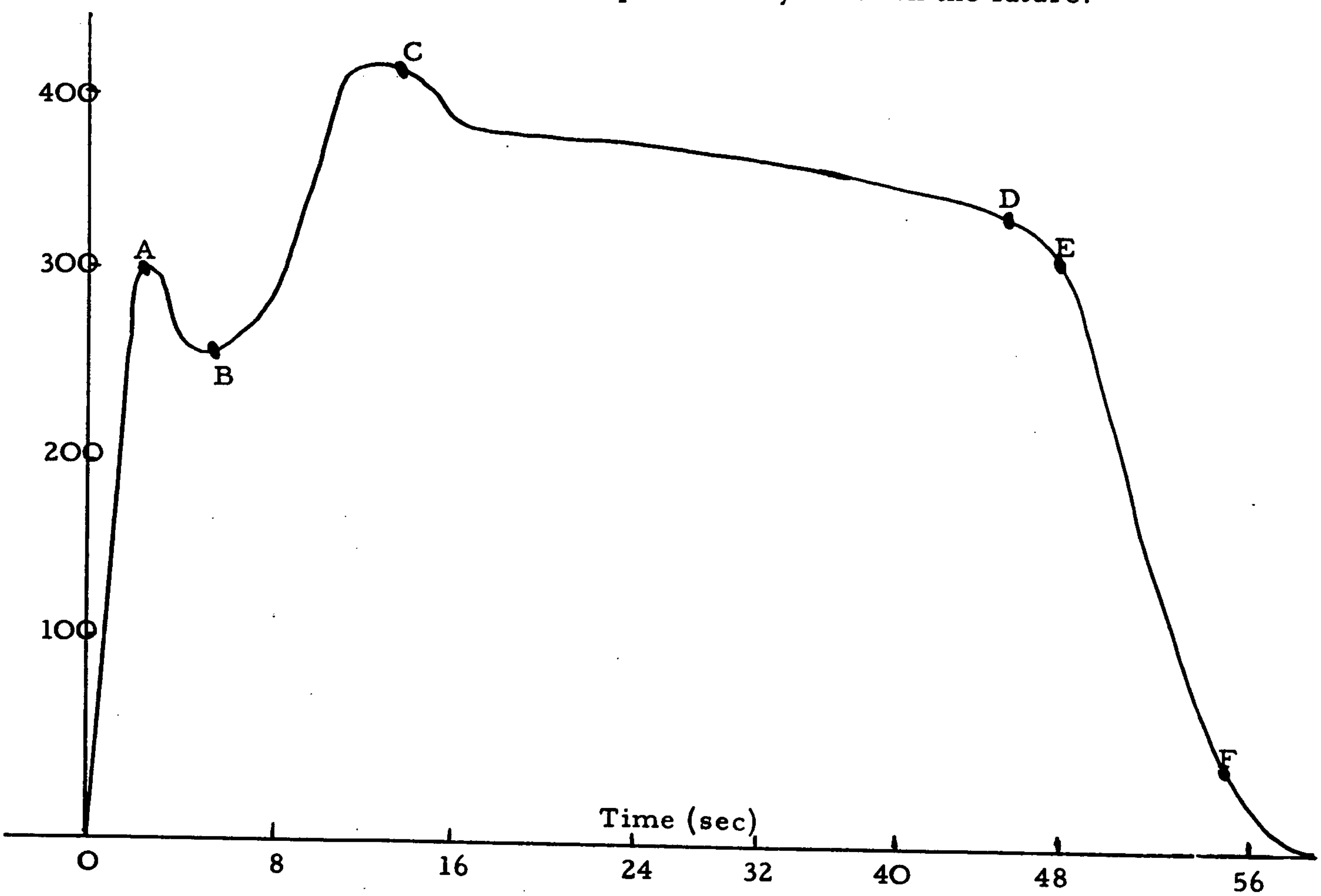


Figure 1

A Typical Thrust Curve for This Study

B. The Second assumption we will make is that the propellant mix from whence the motors are selected as well as the preconditioning temperature can influence the shape of the thrust curve. We will use motors selected from the three propellant mixes (A, B, and C), and conditioned

at three temperatures (0°, 50°, and 100° F). We will use 27 motors, nine randomly selected from each mix, and of the nine, three conditioned at each of the temperatures. This arrangement is illustrated in Figure 2.

Mix	Conditioning Temperature		
	0°	50°	100°
A	3	3	3
B	3	3	3
C	3	3	3

Figure 2

Number of motors selected from each mix and conditioned at each temperature.

C. The third assumption is that temperature conditioning will have an effect similar to that illustrated in Figure 3. These three curves actually represent the average of the curves selected from Mix A and conditioned according to the specified temperature.

D. The fourth and final assumption is that the Test Engineer will want the following information:

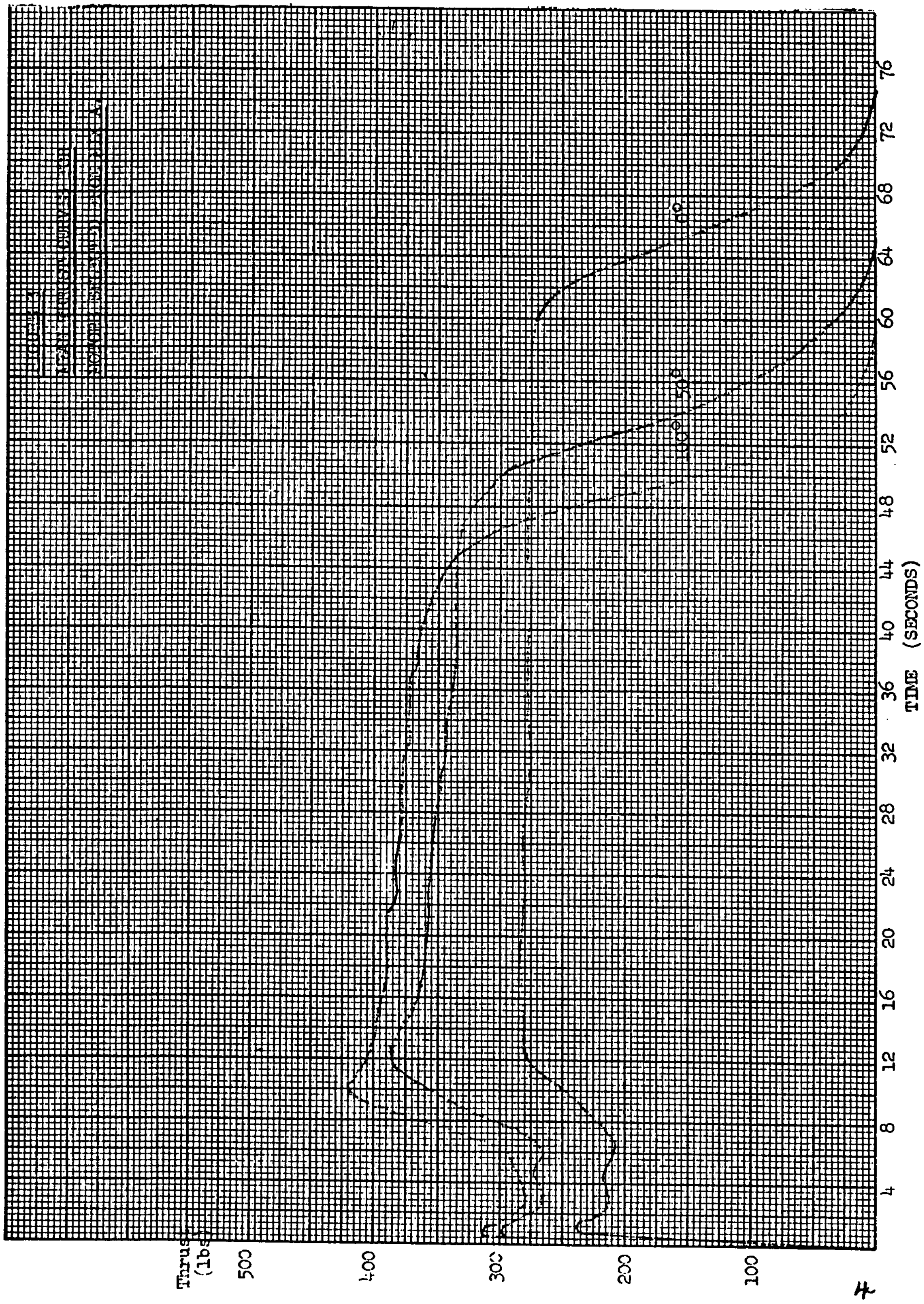
1. Does Mix difference or temperature conditioning have a significant effect upon the shape of a thrust curve?

2. For a given propellant mix or temperature what is the mean or expected thrust curve?

3. In addition to the mean thrust curve, confidence and tolerance bounds are desired.

III. <u>ANALYSIS OF VARIANCE.</u>

A. If you refer again to Figure 1, you will observe a slightly declining plateau between points C and D. I resolved, first of all, to compare


Thrus
(lbs
500
400
300
200
100
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
TIME (SECONDS)
10°
5°
0°
4

performance to this plateau area. One problem is evident from Figure 2, namely the plateau areas are of different length at different temperatures. I therefore decided to study only the region from 15 to 42 seconds inclusive. For all 27 rounds, I read the values at 3 second intervals, that is to say (15, 18, ... 42 seconds). The variances appeared to be homogeneous in the region, so we performed the analysis of variance illustrated in Fig. 3.

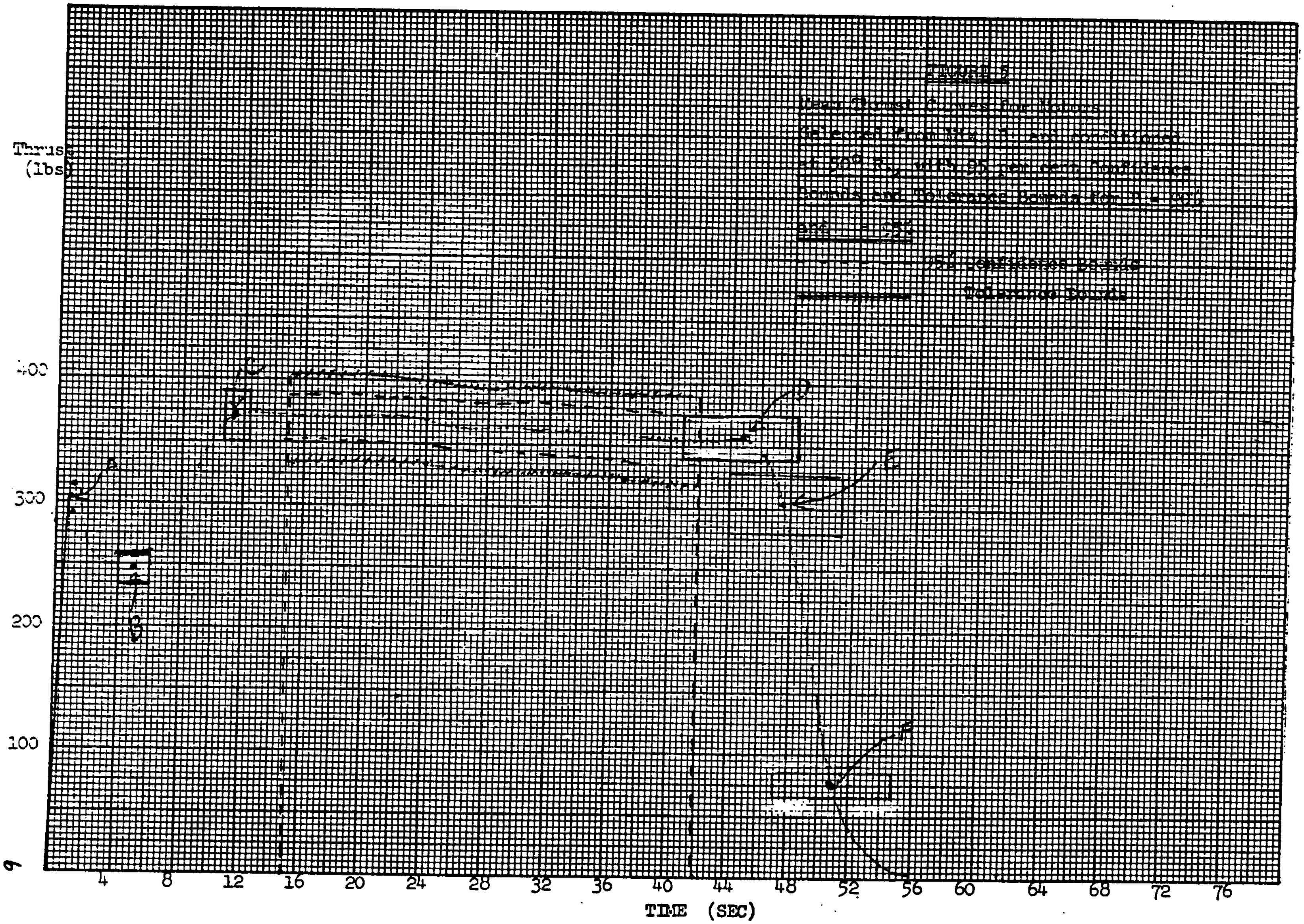
Source of Variance	d/f	SS	MS	F
Mix	2	43217.71	21608..89	82.72**
Temperature	2	488387.22	244193.61	934.78**
Time	9	12325.61	1369.50	5.24**
(Mix)(Temperature)	4	19794.61	4948.65	18.94**
(Mix)(Time)	18	513.70	28.52	0.11
(Time)(Temperature)	18	2954.70	164.15	0.63
(Mix)(Time)(Temperature)	36	24459.88	679.44	2.60**
Error	180	47022.04	261.23 $=(16.16)^2$	
Total	269	638674.97		

*Significant at .05 level (Single and double asterisk used in later tables
**Significant at .01 level have the same meaning indicated here.)

Figure 4

Analysis of Variance of Thrust Values between 15 and 42 seconds for 27 motors

From Figure 4 it appears that for these curves, temperature has a highly significant effect, mix has an important effect, and there is a significant, downward trend during this time interval. While the mix-temperature interaction is significant, its F value is relatively small; so we will assume it is not really critical. It may also be observed that the pooled estimate of variance is 261.23, and we will therefore assume a standard error of 16 lbs. with 180 degrees of freedom. I then proceeded to construct 9 graphs, one for each combination of temperature and mix. Figure 5 illustrates the graph for Mix B and 50° temperature. I took the sample of three and plotted the mean values for the interval from 15 to 42 seconds. Next, I used the pooled estimate of variance and plotted 95% confidence bounds on this curve, and then on the outside of this, I plotted tolerance bounds, $\gamma = 95\%$, $P = 90\%$.


Thrust (lbs)
400
300
200
100
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
TIME (SEC)
95% Confidence Bounds
Tolerance Bounds

Clearly, the question at this point is that of the propriety of arbitrarily taking a time curve, observing the values at stated intervals (every three seconds in this case) and considering time, along with mix and temperature, (3) as one of the treatments in the analysis variance. It is interesting that the error term has 180 degrees of freedom. Had we arbitrarily chosen 2 second intervals, for example, instead of 3 second intervals for taking our readings the degrees of freedom for error would have increased to 270. There is clearly something illogical at this point.

B. Referring again to Figure 1, you will note that I have arbitrarily selected six critical points. I then proceeded by performing an analysis of variance for both the X and Y component for each of these critical components. The Analysis of Variance for the Y component of A is given in Fig. 6 and the X component in Figure 7.

Sources of Var	SS	d/f	MS	F
Mix	198	2	99	0.74
Temp	27,746	2	13,873	104.30
Interaction	200	4	50	0.38
Error	2,398	18	133	
Total	30,542	26		

Figure 6

Analysis of Variance for the Y (Thrust in lbs) Component at Point A.

Sources of Var	SS	d/f	MS	F
Mix	.1267	2	.0634	1.80
Temp	.8339	2	.4170	11.88**
Interaction	.3129	4	.1564	4.48*
Error	.6317	18	.0351	
Totals	1.9052	26		

Figure 7

Analysis of Variance for the X (Time in seconds) Component for Point A

From this it may be observed that temperature had a significant effect but mix did not. Returning to Figure 5, the mean value was located for Point A and a confidence rectangle was drawn about it. This procedure was also followed for the other five critical points and these points were then connected. For the time component, temperature had a significant effect for all six critical points, mix at points D, E, and F. For the thrust Component, temperature had a significant effect at all points except F and mix had a significant effect at points B, D, and E.

One will obviously be concerned at this point by the fact that the six critical points were arbitrarily chosen and are not precisely defined. This could easily result in considerable inaccuracy in collecting data for these points. However, this fact will not necessarily be emphasized since it is not really relevant to the basic purpose of this paper.

However, the matter of performing separate analysis of the time and thrust components of each critical point is highly questionable, and I am certain that a procedure applying the bivariate normal distribution would be in order.

Referring either to Figures 1 or 5, it would appear reasonable that if an analysis of variance is appropriate for C-D, then it would probably be equally appropriate for A-C and possibly for E-F. In fact, it would appear more sensible than attempting to locate and evaluate critical points.

IV. <u>REGRESSION ANALYSIS</u>. A second approach, and one which I feel offers more promise is in the area of regression analysis and polynomial fitting. I will discuss a few ideas along this line at the present time.

If you will refer to Figure 1 again, I arbitrarily broke the graph up into four distinct segments. These are: O-A; A-C; C-D; and E-F. Then, for all 27 motors, I fitted the most appropriate polynomial, that is to say, I fitted a cubic to A-C and straight lines to the other three segments.

I will discuss the procedures I followed in analyzing segment C-D and state little more than that analyses were performed on the other segments, and upon completion all segments were plotted until they intersected. Using the values from 15 to 45 seconds and recording the data at 3 second intervals, a straight line was fitted for all 27 sets of data and an analysis of variance was performed for a, b, and r in the equation $Y = a + b. (X-30)$. Figure 8 gives the analysis of variance and mean values for a, while Figure 9 gives the analysis of variance and means values for b, (note that mix had

no significant effect upon b, so the mean values reflect only temperature). (6) Neither mix nor temperature had any significant effect upon r (the correlation coefficient), but the mean value of the 27 correlation coefficients was 82%.

Sources of Variation	SS	d/f	MS	F
Mix	4411	2	2206	6.28**
Temperature	45959	2	22980	65.42**
Interaction	1959	4	488	1.39
Error	6323	18	351	
Total	58644	26		

Mix	Temperature		
	0°	50°	100°
A	280	348	375
B	292	365	385
C	264	309	376
Ave	278	340	379

Figure 8

The Analysis of Variance and the Mean Values for a, when Fitting the Equation, $Y = a + b \cdot (X-30)$ 15 sec $\leq x \leq$ 45 sec (all means computed from a sample of 3, y = thrust in lbs., x = time in seconds).

Sources of Variation	SS	d/f	MS	F
Mix	0.618	2	0.319	1.040
Temperature	3.689	2	1.844	6.209**
Interaction	0.967	4	0.242	.815
Error	5.340	18	0.297	
Total	10.644	26		

Temperature	0^{o}	50^{o}	100^{o}
Mean Value b	-.410	-.881	-1.315

Figure 9

The Analysis of Variance and Mean Values for b, obtained from fitting the equation $Y = a + b\cdot(X-30)$, 15 sec. $\leq x \leq 45$ sec. (all means computed from a sample of 9)

The data in Figure 9 indicates that "b" increases almost linearily with temperature. In fact, the formula $b = -.410 - (.00905)$ temp, might serve as a guide for selecting "b" in the region $0^{o} \leq \text{temp} \leq 100^{o}$. If one desires a formula for estimating "a" in the region for $0^{o} \leq \text{temp} \leq 100^{o}$, he might try the formula: $a = 278 + 1.4 \cdot (\text{temp}) - .0041(\text{temp})^2$ which averages out the mix effect.

Again referring to Figures 8 and 9, one may observe that the standard error for "a" is 18.7 lbs. with 18 degrees of freedom, and the standard error for "b" is 0.545, also with 18 degrees of freedom.

In addition to this, each time a line is fitted by least squares, it is possible to obtain a standard error of estimates and standard errors for "a" and "b". For the 27 curves, I pooled these standard errors and obtain the following results:

Pooled Standard Errors of estimates:
3.40 lbs. with 243 d/f.

Pooled Standard Error for "b":
0.324 with 243 d/f.

Pooled Standard Error for "a":
1.025 with 243 d/f.

As may be expected, the estimates for the standard error for both "a" and "b" are larger in Figures 8 and 9 than the pooled estimates listed above. This is reasonable since the estimates in Figures 8 and 9 include the dispersions that exist among curves from the same lot and conditioned at the same temperature, while the pooled estimates reflect the variation within only a single curve.

Inasmuch as I am attempting to classify any curves which come from a given mix and a given temperature, it would seem more appropriate to use the estimates of variability in Figures 8 and 9. One other argument for this lies in the fact that when the standard error of estimates was computed for each of the 27 curves, it was computed from 11 points, selected from the thrust curves at 3 second intervals. 15 sec $\leq$ time $\leq$ 45 sec. Again the question arises concerning the arbitrariness in choosing 3 second intervals instead of some other intervals.

Now confidence bounds for a regression line at a point x_i may be computed from the formula

$$\bar{Y}_i - t \cdot S(\bar{Y}_i) \leq E(Y_i) \leq \bar{Y}_i + t \cdot S(\bar{Y}_i)$$

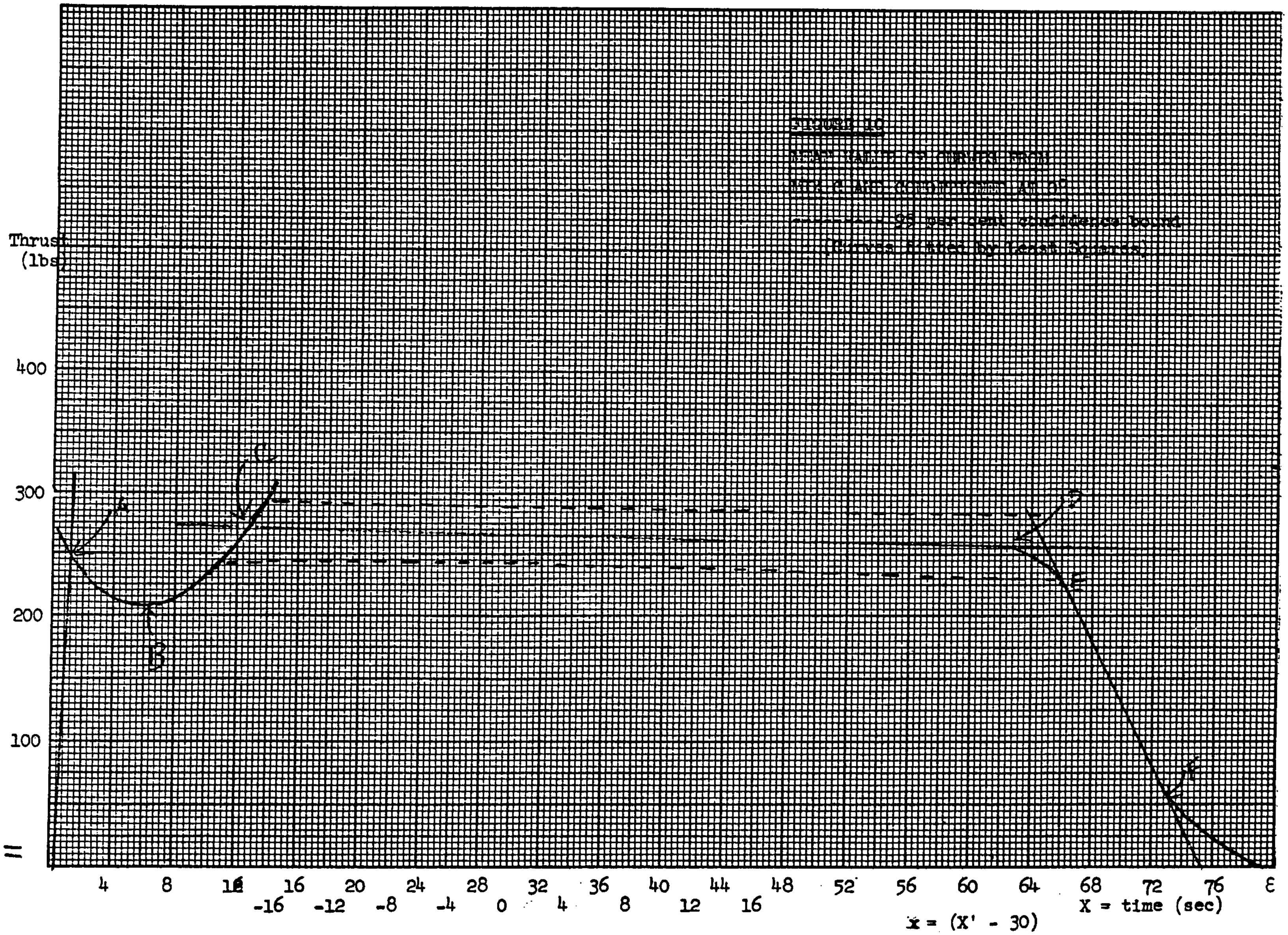
where $\bar{Y}_i = a + b\, x_i$

and $S^2(\bar{Y}_i) = S^2(a) + x_i^2 \cdot S^2(b);\ \bar{x}_i = 0.$

Since each "a" represents an average of 3 numbers and each "b" represents an average of 9 numbers, we have:

$$S^2(\bar{Y}_i) = \frac{351}{3} + x_i^2 \cdot \frac{.297}{9} = 117 + .033\, x_i^2\ .$$

These ideas are illustrated in Figure 10, in which we consider Mix C and a temperature of 0^o, and fit the curve from 15 sec $\leq X \leq$ 45 sec. (Segment C-D). This curve is given by $\bar{Y} = 264 - .410\, X$. or $\bar{Y}=264-.410\cdot(X-30)$,


95 per cent confidence bound
(Curves fitted by Least Squares)
Thrust (lbs)
400
300
200
100
4
8
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76
-16
-12
-8
-4
0
4
8
12
16
X = time (sec)
x = (X' - 30)
A
B
C
D
E
F

and the variance $S^2(\bar{Y}) = 117 + .033 X_i^2$ was used to compute a confidence bound about the curve. Incidentally, the confidence bound in Figure 10 is very close to the one illustrated in Figure 5.

The procedures for fitting the segments (O-A) and (E-F) could be quite similar to that of fitting the segment (C-D). In fact the segment (O-A) should be even simpler. To fit the segment (A-C), it is suggested that a cubic equation be fitted, using data points at one second intervals. It is further suggested that orthogonal polynomials be used when fitting a cubic or higher degree equation to simplify the process of obtaining the variance (7) and confidence bounds.

CONCLUSIONS. Frequently it is desired to design an experiment when the results of the test are a continuous curve rather than a single quantitative value. Scientists and Engineers frequently want to know whether certain levels of a given treatment will have a significant effect upon the curve obtained, and what will an average or expected curve be for a given set of conditions.

I have made a few suggestions based largely upon analysis of variance or regression analysis. I will greatly appreciate comments on the proposed solutions, but more important, I would like suggestions for better approaches to the problem.

QUESTIONS.

1. Has the problem of designing an experiment when the results come as a continuous curve rather than a single value ever been solved? If so, are useful references available?

2. Do you have any ideas of additional approaches beside those suggested in the paper?

3. When studying a section of the curve such as C-D, is there any justification in arbitrarily selecting a set of times between C and D, computing the thrust at each of these times for all available curvex, and performing an analysis of variance similar to that given by Figure 4? Perhaps this would be in order with certain changes in procedure.

4. What procedures would you suggest when attempting to locate a point in terms of both its X and Y components and then obtaining both a confidence and tolerance region about this point?

5. Is the analysis of variance for a, b, and r, as illustrated in Figures 8 and 9 appropriate?

6. Have you any suggestions regarding the validity of the techniques, using regression analysis, that were discussed in this section?

COMMENTS ON PRESENTATION BY PAUL COX

Walter T. Federer
Mathematics Research Center, United States Army
University of Wisconsin, Madison, Wisconsin

The paper presented by Mr. Cox is written in a somewhat provocative manner. I appreciate this style of presentation as it affords the Panel ample opportunity to illustrate several statistical points.

The first point I wish to make relates to the definition and use of terms in current statistical literature. There is a tendency in statistical literature for vague and imprecise usage of such terms as the design of experiments, analysis of variance, error rate, etc. It is instructive and useful to define and to use words or phrases in a specified manner. Any departure from specificity should be described. Personally, I would prefer to use definitions of the following form:

i) Experimental design (or experiment design) - The arrangement of the observations in the experimental area or space or the procedure for obtaining the observations in an experiment.

ii) Treatment design - The arrangement or selection of treatments for the experiment (e.g., the selection of levels and combinations of factors in factorial experiments, etc.)

iii) Determination of sample size - The number of observations necessary to achieve a prescribed objective. (Authors of some ranking procedures papers refer to the determination of numbers of observations as the design of experiment rather than as the determination of sample size.)

iv) Analysis of variance - The partitioning of the sum of squares into component parts. (One segment of statistical literature utilizes the term analysis of variance to be synonymous with an F test while another segment utilizes this term to refer to the estimation of variance components and so it goes.)

v) Analysis of experimental data - This term includes the last above but not vice versa. It refers to all statistical computations relevant to a set of experimental data. An analysis of experimental data refers to the reduction of data to summary form and is useful in, but does not replace, the interpretation of experimental results. The interpretation of statistical results must be made in light of the objectives, conditions, and related circumstances of the experimental results.

vi) Significance level - Type I error = size of the test = α, have all been used to refer to the same thing but unfortunately nothing is said about the base for computing "α".

vii) Valid estimate of the error variance - Fisher has defined this term but unfortunately many statistical writers by-pass this important concept with the phrase "given that σ^2 is the error variance." In much of experimentation the definition of error variance cannot be so glibly by-passed, but requires a thorough knowledge of the experimental conditions.

We could go on with other terms but now let us return to Mr. Cox's paper. The title of the paper is "Statistical Design of Experiment for Continuous Data"; it deals only with the analysis of experimental results with no reference either to the experimental or treatment design as defined above. Mr. Hartley has discussed some considerations to be given to the treatment design for experiments with specified objectives. Mr. Lucas will, I hope, make some comments about the actual experimental design used in this study and illustrate where confounding has taken place. Mr. Cox's paper is concerned with what to do with a set of data and not with how to obtain the data. He has raised a number of questions but rather than address myself to the specific question I prefer to proceed in another manner which, I hope, will furnish answers to or illustrate the relevance of the questions.

As Messrs. Grubbs, Greenberg, Hartley, and Schneiderman have already stressed we must first set up a Mathematical Model for the data which will be consistent with the experimental and treatment designs and with the nature and objectives of the experiment. For example, let us suppose that thrust = y, may be characterized by the following:

$$y = f(\underline{\epsilon}, \underline{t}, \underline{\theta})$$

where the response variable y is a function of error components denoted by the vector $\underline{\epsilon}$, of time components denoted by the vector $\underline{t}$, and of a set of parameters denoted by the vector $\underline{\theta}$. Our first job then is to define to nature of the function. If we are totally ignorant of the response curve then we could use a form of polynomial regression as follows:

$$E(y) = \sum_{i=o}^{b} \beta_i t^i$$

where β_i is the i^{th} regression coefficient and t the time variable. After we are satisfied that a suitable mathematical formulation of the problem has been made, the parameters of the response curve are estimated. The analysis of the estimates may be made using the results of R. A. Fisher (Jour. Agric. Sci. 11:107, 1921 and Phil. Trans. Roy. Soc. B, 213:89, 1925) and others. Also, multivariate analysis procedures may be pursued for summarizing the results for many estimates of a set of parameters. For example, if it is desired to discriminate between response curves, then an a priori or an a posteriori (These terms are not reserved solely for use by Bayesians.) weighting of coeffients in the discriminate function may be utilized.

As a part of the characterization of the model and of the problem it should be determined if the total response curve segements of the total curve, or specified points (e.g. points of inflection) on the curve are of interest. After this has been specified then the statistician proceeds with the estimation problems. Haziness on form or type of response desired leads to a confusion of issues.

One specific question raised by Mr. Cox related to the sample size N for response curves for continuous data. Now if the data are truly continuous N = infinity, but we all know that the recording machine records an impulse over a measurable period of time, say one-tenth of a second. In any event N is very large. Several of the previous Panel speakers have discussed the non-independence of two successive impulses or recordings by a recording machine. However, I wonder about the relevance of this since we use, or should use, these values only to estimate the parameters in the response curve. This procedure is, or should be, repeated for many response curves and the variation among response curves treated alike forms a basis for the variances and covariances among the estimates of parameters where each response curve represents but one observation.

At this point I do not see the importance of obtaining a variance of a single response curve. However, if such is desired, then as an approximation I would suggest segmentation of the total curves into small segments of time where small is such that the estimates are relatively unaffected by smaller segmentation. Course groupings could affect the results considerably. Some account may need to be taken of the relationship among adjoining segments as described by Messrs. Greenberg and Hartley.

The response curves presented in the paper bother me somewhat. Frankly, I believe (i) that the curves in Figure 3 are not very fictitious, (ii) that the area under each curve is relatively constant from the conservation of mass theory, (iii) that a heart-to-heart talk with the physicists and engineers would do much to simplify the nature of the problem, and (iv) that maybe Mr. Cox should be considering acceleration = z instead of thrust = y.

Summed up this means that I would want some education in this area before any analyses would be performed on thrust or any other data. It may be possible to reparameterize the problem by using a function of the time variable instead of the time variable itself. Some simple function such as log t might suffice.

COMMENTS ON PRESENTATION BY PAUL COX

Bernard G. Greenberg
School of Public Health
The University of North Carolina
Chapel Hill, N. C.

As the first discussant, my main concern will be with what I consider the the most important aspect of this problem. It involves the question of the basic underlying model and what the purposes of the investigator were in designing this experiment. As our chairman has just pointed out, this is our primary consideration.

This example illustrates what I think is a truism in statistical design and experimentation, viz. that no amount of statistical knowledge and methodology, however great, is a sufficient substitute of substantive knowlege of the field of application and in being able to discuss with the investigator the important questions. For instance, Mr. Cox has indicated that his six points were arbitrarily selected and the analysis of variances relating thereto were not necessarily meaningful. This may be so in these six points but does the person who has the expert knowledge of thrust curves tend to associate important meanings to them, or any other set of critical points. It may very well be that point A has direct application to the understanding of the model underlying thrust curves and that the actual value of A, the time elapsed from the origin, or the rate of ascent from the origin to A are of primary concern to the rocket motor designer.

This is best illustrated from examples in my own field of biology and medicine. We have similar kinds of data which may not be as closely continuous as the thrust curve, but the points of observation in the time series are spaced close enough together that we treat them as such. Thus, we have tracings for electrocardiograms, growth curves, and epidemic curves. In the electrocardiogram, the distance and regularity between two waves is extremely important to the cardiologist and deviations from normalcy are based upon this pattern. In an epidemic curve, we are interested in the length of time between peaks such as in the periodicity of measles. We are also interested in the amplitude of the waves in order to know expected numbers of cases.

In other words, without discussing with the engineer of the rocket motor thrust curve what his problem is, it is difficult for me to know how to handle best these data. It may be the measurement of the critical points, the length of time until each maximum is attained, or even the integral summing up the total amount of thrust.

If one is convinced that he does want to fit a mathematical equation to this curve, or segments of it, a nasty problem arises because of the time series nature of the data. The observations in time are not independent and the residuals may be autocorrelated. This is not a new problem. The econometricians have been considering this problem for years in their analysis of time series.

I might suggest one reference to Mr. Cox in this connection which may help in fitting equations to segments of the curve and the analysis thereof. A paper by Elston and Grizzle (Biometrics, Vol. 18, No. 2, June 1962, pp. 148-159.) considers the various ways of estimating time-response curves and the ideas in that paper should prove helpful in this problem.

COMMENTS ON PRESENTATION BY PAUL COX

H. L. Lucas
Institute of Statistics
North Carolina State College
Raleigh, North Carolina

The panelists so far have covered just about everything that I had in mind. I certainly agree in the main with the comments they have made regarding just what particular points on the observed curve or what particular function of the observations may be of interest. Also, I agree with the comments regarding the desirability of fitting a "rational" model, which presumably can be supplied, at least in approximate form, by the engineers. I wish, however to expand on a very important point.

Many of the remarks of panelists about design and analysis have been engendered by the existence of "noise" along the curve for an individual motor and the probable lack of independence of successive observations. I wish to emphasize that there is another, and probably much more important, "noise" component involved. The latter arises from the fact that a group of motors which are constructed and treated alike, insofar as can be managed, will, nevertheless, have inherently somewhat different curves. That is, there is "between-motor" noise as well as "within-motor" (along-the-curve) noise. The existence of between-motor noise must be taken into account for proper experiment design and analysis.

It is instructive to formalize the situation in a way which encompases the two noise components. For the jth experimental unit (here the motor, but in other cases a machine or an animal, etc.) on the ith treatment, we can write the model,

$$y_{ij}(t) = \phi(t; \underline{\theta}_{ij}) + \epsilon_{ij}(t) \tag{1}$$

where

$y_{ij}(t)$ = observed time curve for the unit

$\phi(t;\underline{\theta}_{ij})$ = "true" time curve for the unit

$\underline{\theta}_{ij}$ = vector of parameters for the unit

$\epsilon_{ij}(t)$ = "within-unit" noise,

For the jth unit on the ith treatment, we next write

$$\underline{\theta}_{ij} = \underline{\theta}_i^* + \underline{\delta}_{ij} \tag{2}$$

where

$\underline{\theta}_i^*$ = expected value of $\underline{\theta}_{ij}$ for units on the ith treatment
$\underline{\delta}_{ij}$ = "between-unit" noise.

Substituting (2) into (1) yields the model desired, namely,

$$y_{ij}(t) = \phi\left[t; (\underline{\theta}_i^* + \underline{\delta}_{ij})\right] + \epsilon_{ij}(t). \tag{3}$$

Suppose we compute $\underline{\theta}_{ij}$, an estimate of $\underline{\theta}_{ij}$, for each unit. We see that

$$\hat{\underline{\theta}}_{ij} = \underline{\theta}_{ij} + \underline{\eta}_{ij} \tag{4}$$

where

$\underline{\eta}_{ij} = \underline{\eta}\left[t; \underline{\theta}_{ij}; \epsilon_{ij}(t)\right]$, a vector of errors with which $\underline{\theta}_{ij}$ is estimated; these stem from "within-unit" noise.

We are interested, however, in estimating $\underline{\theta}_i^*$. The relation of $\hat{\underline{\theta}}_{ij}$ to $\underline{\theta}_i^*$ can be seen by substituting (2) into (4) to obtain

$$\begin{aligned} \hat{\underline{\theta}}_{ij} &= \underline{\theta}_i^* + \underline{\delta}_{ij} + \underline{\eta}_{ij} \\ &= \underline{\theta}_i^* + \underline{\delta}_{ij}\,. \end{aligned} \tag{5}$$

Note that $\underline{\delta}_{ij}^* = \underline{\delta}_{ij} + \underline{\eta}_{ij}$ is the total noise or error in $\hat{\underline{\theta}}_{ij}$ and stems from both "between" and "within" noise.

In view of the development just completed, it is certainly reasonable first to estimate $\underline{\theta}_{ij}$ for each individual unit and then as a second step, to analyze the $\hat{\underline{\theta}}_{ij}$ according as the experimental design dictates. Since $\hat{\underline{\theta}}_{ij}$ is a vector, multivariate methods may be desired. Note that the procedure is a "robust" one.

Some papers in which the "robust" approach has been employed are [1], [3], [4], [5], [6].

In view of the remarks of some of the other panelists about choice of points along the time curve and about correlation between successive observations along the curve, the following comments seem in order. In my experience, the contribution of the "between" noise, $\underline{\delta}_{ij}$, to the variance of $\hat{\underline{\theta}}_{ij}$ as an estimate of $\underline{\theta}_i^*$ is dominant over the contribution of the "within" noise as summed up in $\underline{\eta}_{ij}$. In fact, in some instances, the "between" noise, $\underline{\delta}_{ij}$, is large relative to the "within" noise, $\epsilon_{ij}(t)$, itself; in this event, the contribution of $\underline{\eta}_{ij}$ is negligible. With $\underline{\delta}_{ij}$ dominant over $\underline{\eta}_{ij}$, it is clear that one need not worry much about the correlation between successive observations on the same unit, that any reasonable method of computing $\hat{\underline{\theta}}_{ij}$ will do, and that one needs use only the minimum number of points along the ij^{th} curve consistent with the complexity of ϕ and the obtaining of moderately efficient estimates of $\underline{\theta}_{ij}$.

This leads next to the design problem, a matter which has been discussed by the other panelists primarily from the standpoint of selecting points along the time curve. In view of my foregoing remarks, I cannot see that the pattern for selection of points along the time curve is the really critical matter, just as long as the pattern is a reasonable one. Instead, the important question is how to select an optimum set of treatment combinations.

To comment further about the design problem, it is again advantageous to be somewhat formal. We note that $\underline{\theta}_i^*$ is a function of the levels of the

treatment variables (here, temperature and mixture); i.e.,

$$(6) \qquad \underline{\theta}_i^* = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha})$$

where

γ = a vector of functions of the vectors $\underline{x}_i$ and $\underline{\alpha}$

$\underline{x}_i$ = the vector of levels of the treatment variables characterizing the i^{th} treatment;

$\underline{\alpha}$ = a vector of parameters which depends on basic invariants and on the levels maintained for treatment-type factors not under study (i.e., factors held constant over all i).

Substituting (6) into (5) yields

$$(7) \qquad \hat{\underline{\theta}}_{ij} = \underline{\gamma}(\underline{x}_i;\ \underline{\alpha}) + \underline{\delta}_{ij}^* .$$

Now, if the functional forms represented by $\underline{\gamma}$ are known, the problem is to select a minimum optimal set of x-vectors such that all elements of $\underline{\alpha}$ can be estimated and that the estimate, $\hat{\underline{\alpha}}$, is "best" in a suitable sense. In general the optimum design depends on $\underline{\alpha}$, but, since $\underline{\alpha}$ is unknown, one must use previous estimates (or best guesses) about $\underline{\alpha}$ in order to arrive at a good design. Some ideas about this problem are given in [2]. If the forms of the functions, $\underline{\gamma}$, are subject to question, the design must have extra x-vectors so that tests about the assumed $\underline{\gamma}$ and insight about improvements can be obtained. The latter point is also discussed briefly in [2].

I have finished the main things I want to say. There are, however, a couple of other matters that come to mind.

The first has to do essentially with what function of ϕ and hence of $y_{ij}(t)$ is really of concern to the investigator. Although, in some instances, only a particular univariate function of ϕ may ever be of interest, my experience indicates that this is not generally true. I suggest, therefore, that ordinarily it will be best to study ϕ; i.e., to fit the parameters, $\underline{\theta}_i^*$,

or more basically, $\underline{\alpha}$. Given such fits, anything desired can be ascertained.

Finally, in the first analysis Mr. Cox outlined, he failed to distinguish "between" and "within" noise. The variance sources for his analysis were

Treatment

Time

Time by treatment

Residual.

They should have been

Treatment

Motor within treatment (Error for treatment; corresponds to δ^*_{ij}

Time

Time by treatment

Time by motor within treatment (Error for time and time by treatment; coresponds to ϵ_{ij}).

In closing, I should note that Mr. Cox, in all but his first analysis, adopted the "robust" approach. I stress the approach, however, because it is important, and because judging from his first analysis, Mr. Cox appeared not to be very clear on the implications of the existence of both "between" and "within" noise.

REFERENCES

[1] Box, G.E.P. 1950. Problems in the Analysis of Growth and Wear Curves. Biometrics 6, No. 4, 362-389.

[2] Box, G.E.P. and H.L. Lucas. 1959. Design of experiments in non-linear situations. Biometrika 46: 77-90.

[3] Elston, R.C. and J.E. Grizzle. 1962. Estimation of Time-Response Curves and Their Confidence Bands. Biometrics 18, 2, 148-159.

[4] Fisher, R.A. 1920. Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain from Broadbalk. Jour. Agric. Sci. 11: 107-135.

[5] Fisher, R.A. 1924. The Influence of Rainfall on the Yield of Wheat at Rothamsted. Phil. Trans. Roy. Soc. B, 213. 89-

[6] Vanasse, N.A., I. D. Jones and H. L. Lucas. 1951. Specific gravity-dry matter relationship in potatoes. Am Potato Jour. 28: 781-791.

COMMENTS ON PRESENTATION BY PAUL COX

H. O. Hartley
Statistical Laboratory and Department of Statistics
Iowa State University
Ames, Iowa

As the last of the discussants to give my comments in writing I have the advantage of a preview of what my predecessors have said. I will attempt to summarize and amplify their competent comments.

Mr. Cox has certainly described a problem which has raised many questions of considerable interest. His problem is concerned with the design and analysis of an experiment in which the 'response' is measured in the form of a curve (thrust curve of a motor is his example). Actually Mr. Cox is almost exclusively concerned with analysis. I will later make a few remarks on the design aspect.

Let me start by saying that we cannot really talk about an 'appropriate analysis' of a set of experimental response curves without being clear about

(a) The purpose or the objectives of the experiment

and at least to some extent about

(b) The physical mechanism generating the experimental responses.

With regard to (a) Mr. Cox has described essentially two objectives, namely, the effect of 'Conditioning Temperature' and 'Mix' on the 'Shape of the Thrust Curve' and the 'Total Impulse'. The latter is a single response clearly defined as an integral of the response curve and obtainable (say) by numerical integration. The latter requires clearer definition in terms of thrust curves characteristic of real interest to the engineer and to be specified by him. We may speculate that one of these may be the initial rate at which the thrust increases from zero, or possible the time at which it reaches the stationary stage, etc. Mr. Cox has, however, pointed to an important feature, namely that in general a multiplicity of responses will have to be computed from the curve representing the relevant summaries of interest to the engineer. Following Dr. Lucas's notation and denoting by $y_{ij}(t)$ the thrust for the j^{th} unit of the i^{th} treatment group observed at time t, we would compute for each curve k summaries $S_r(y_{ij}(t))$; $r = 1, 2, \ldots, k$, which in this example, may well be computed from standard formulas of numerical integration and differentiation. To answer the purpose of the experiment we may in many cases apply the well established techniques of multivariate (k-variate) analysis of variance (see, e.g., Smith, H; Gnanadesikan, R. and Hughes, J. B. (1962)) to the $S_r(y_{ij}(t))$ which would be a 3×3 factorial (i) by temperature

and mix with 3 replicate units (j) in each cell. Much of the information will often be obtainable from a standard single variate analysis of variance applied to each of the $S_r(y_{ij}(t))$ separately. This sort of analysis which uses a 'between unit error' is also recommended by Dr. Lucas and Dr. Federer, called 'robust' by the former, and is in essence identical with Mr. Cox's analysis of variance for the regression intercept, a, and slope, b, fitted to a 'straightlooking' section of the curve. I question, however, whether Mr. Cox's procedure of 'arbitrarily' breaking up the curve into sections and fitting polynomials separately to the sections really contributes to our appreciation of the engineering aspects Is it really of interest to the engineer that a cubic term in the first section goes up with temperature? To my mind it is of the greatest importance to communicate with the engineer on the selection of relevant summaries $S_r(y_{ij}(t))$.

This brings me to (b), namely, the importance of a physical theory leading to a mathematical model for the thrust $y_{ij}(t)$, stressed by all discussants. Dr. Lucas postulates a model of the form

$$(1) \qquad y_{ij}(t) = \phi(t; \theta_{ij}) + \epsilon_{ij}(t)$$

where θ_{ij} is a (say) m vector of parameters. Whilst in the present example it should be quite feasible to obtain such a model from (say) the differential equations governing the dynamics of the thrust phenomenon, the statistician may be called upon to analyze curves arising in a situation in which the setting up of a mathematical model is difficult. I would stress, therefore, that summaries $S_r(y_{ij}(t))$ answering the purpose of the experiment can often be decided upon without reference to a mathematical model, although the study of their statistical efficiency is facilitated by the mode. Where the latter is available one may proceed as Dr. Lucas suggests to estimate the 'treatment averages' θ_i^* of the θ_{ij} although it may be argued to be more appropriate to estimate the treatment averages of the relevant summaries $S_r(\phi(t; \theta_{ij})$ the two being differentially equivalent. Whatever method is used I believe that some attention should be given to the estimation of the individual units', θ_{ij}, and this raises the question of the 'within curve' error or noise. I agree with Dr. Lucas that this will often be

relatively unimportant. However, it is of the same degree of relevance as, for example, the estimation of the mean life for a time mortality curve or the L. D. 50 of a dosage mortality situation both of which are usually 'within curve' estimation problems and both have received considerable (possible exaggerated) attention by statisticians. I will, therefore, answer the question raised by Mr. Cox concerning the within curve error structure: --The 'degrees of freedom' that are to be attached to a set of risiduals $y_{ij}(t_s) - \phi(t_s, \theta_{ij})$ computed at some arbitrarily selected time points, t_s, appear to depend on the choice of the t_s. Because of the time series correlogram the sum of squares of residuals

$$\sum_{s=1}^{s} (y_{ij}(t_s) - \phi(t_s, \theta_{ij}))^2$$

is approximately distributed as $c\chi^2$ based on an 'effective number' of degrees of freedom (see, e.g., Bayley and Hammersley (1946)) and ideally this should be invariant with the choice of grid points, t_s. Without the knowledge of the correlogram one cannot judge how the degrees of freedom of Mr. Cox's Figure 4 are affected. However, the analysis given in this figure in any case does not take proper account of the distinction between 'within curve' and 'between curve errors' as was pointed out by Dr. Lucas.

Finally a few words on the question of the design of an experiment with curve responses. First let me say that the choice of t_s values is not a question of the design (as Dr. Lucas rightly stresses but as far as I recall nobody said so during the discussion). This is a computational question of analysis. The design is concerned with the choice of the levels of 'temperature' and the composition of the mixes or, indeed, with questions of what treatment combinations should be chosen. Since the work by Box and Draper (1959) and Kiefer (1959) and others is mainly concerned with a single experimental response or a single response surface, much needs to be done about designing experiments which in some sense are optimum for the 'assessment' of multiple response surfaces, particularly if (as is the case with Mr. Cox's example) some factors (mix) are qualitative. In the absence of a comprehensive theory one would perhaps single out one important response from the $S_r(y_{ij}(t))$ and optimize the design for it using some such theory as the above but optimizing subject to tolerances for the bias and precision with which reponse surfaces for the other response surfaces can be estimated.

REFERENCES

Bayley, C. V. and Hammersley, J. M. (1946). Suppl. J. Roy. Stat. Soc. 8, 184.

Box, G. E. P. and Draper (1959). J. Amer. Stat. Assoc. 54, 622.

Smith, A., Gnanadesikan, R. and Hughes, J. B. (1962). Biometrics, 18, 32.

Kiefer, J. (1959). J. Roy Stat. Soc., Series B, 21, 272.

COMMENTS ON PRESENTATION BY PAUL COX

Marvin A. Schneiderman
Biometry Branch, National Cancer Institute
National School of Health, Bethesda, Maryland

It is always a pleasure (and enlightening) to attend a statistical session in which Professor Greenberg and Professor Hartley participate. They usually make important remarks, succintly, and thus save one from the responsibility of adding much more than "amen." Since they've spoken before me, I wish now to add my "not much more", (as well as "amen").

Mr. Cox has presented us with a very interesting problem in engineering which has a direct counterpart in medical and biological research. I have in mind among other measures the interpretation of electrocardiogram tracings, electro-encephalogram tracings, and the flood of apparently continuous measures that the physiologists are now capable of making. From this analogy, I am led to take slight issue, however with one of Professor Hartley's remarks.

A single response measure, such as the LD_{50} may well be inappropriate. Of course, Mr. Cox's engineer has been rather vague about what he really wants to know, and I suspect that more extended discussion with the statisticians might lead the engineer to specify his problem more. I am guessing that an LD_{50} would be inappropriate. Assume that the statistician finds, though, that he is interested in the shape of the curve, in some sense. That is, the specific shape of the curve, or the presence or absence of some specific wiggle tells him something about the physics underlying the system. For example, in the biological counterpart, a straight line inactivation curve (log response vs. time, for example) as was first postulated for the Salk vaccine for polio implies a simple one-step chemical process, or a single manner of excretion. A concave upward curve may imply a two-step process, or two (or more) modes of excretion, (i. e. a sum of exponentials) or some other functional arrangement. The investigation of the kinetics of such systems are a whole sub-field. Professor Hartley has some good advice to offer in the fitting of these sums of exponentials. Thus, the curves themselves may be the items of most importance to the experimenter. This should not be lost.

On the specific suggestions made by Mr. Cox for the analysis of the data, the physical meanings of the various points on the curve, A, B, C, D, E, F, might help guide the statistician into more fruitful lines--perhaps even into a solution of the problem the engineer wants solved. As Professor Hartley

pointed out, the points on the curve are correlated, making the analysis of variance inappropriate because of the non-independence. One earlier reference on the non-independent regression problem then the Bailey-Hammersley reference that may be helpful is one by John Mandel, in the Journal of the American Statistical Association "Fitting a straight line to certain types of cumulative data," Vol. 52, p. 552 (1957). My recollection is that Mandel shows that a least squares approach gives an unbiased estimate of the parameters, but gives the wrong (too small) variance. He gives other references to this problem, too.

On the "shape" problem, there is a paper by G.E.P. Box and W.A. Hay which may be of interest. It appeared in Biometrics, Vol. 9, p. 304 (1953) "A statistical design for the efficient removal of trends occurring in a comparative experiment, with an application in biological assay." A recent doctoral dissertation by Francis J. Wall, from the University of Minnesota considers an aspect of the nearly continuous data problem in biology and medicine. The title is "Biostatistical linear models in longitudinal medical research problems." The title shouldn't mislead engineers. It's much the same problem as we had here.

THE INDEPENDENT ACTION THEORY OF MORTALITY AS TESTED AT FORT DETRICK

Francis Marion Wadley
U. S. Army Biological Laboratories
Fort Detrick, Frederick, Maryland

The independent action theory is sometimes used as an approach to all-or none dosage-effect problems instead of the more usual dosage-effect methods such as probit analysis. With the probit and similar analyses, the basic assumption is of varying susceptibility among the subjects. With the independent action theory in its simplest form the assumption is that any toxic unit reaching the site of action will be effective. Each unit is believed to have a small but definite chance of hitting its mark; a higher percentage response to larger doses is produced by multiplication of this chance. This theory obviously does not assume varying susceptibility among subjects, and will logically lead to the same slope for all trials. If $\underline{a}$ is the chance of hitting the mark, the chance of escape is (1-a) for one toxic unit; for 2 units, it is $(1-a)^2$; for $\underline{n}$ units $(1-a)^n$. Danger from bullets on a battlefield has been used as one illustration.

The independent action theory was apparently first developed by Neyman and associates, according to K. L. Calder of Fort Detrick. It has been used by Watson (Phil. Trans. Roy. Soc. London, 1936) in studying transmission of plant viruses by insect migrants. It is also used by some workers in dosage-effect studies where the dose is of biological agents (S. Peto, Biometrics 1953; L. J. Goldberg and associates, 5th, 8th and other reports, Navy Biol. Lab., 1951-52). A. W. Kimball (1953 lectures, Fort Detrick) has applied the theory to radioactive particles. Peto presents detailed procedure for calculation, and mathematical methods are also presented by Andrews and Chernoff (Tech. Rept. 17, Applied Math Lab, Stanford, 1952), and by W. G. Cochran (1946 lectures, N. C. State). Goldberg (l. c.) has worked out special plotting paper for quick graphic estimation of LD-50 and its error. The extensive work on dosage theory assuming varying susceptibility is conveniently summarized by Finney. ("Probit Analysis", Cambridge U. Press, 1952).

Where such agents as bullets and radioactive particles are considered, there can be no question that the idea of independent action will apply better than the concept of dosage and varying susceptibility. With chemical toxicants which can be measured out accurately, the

idea of dosage and varying susceptibility undoubtedly applies better than the independent action concept. Susceptibility is know to vary. With biological agents such as pathogenic bacteria, we are on a middle ground where either procedure may have its advocates. The basic question appears to be whether susceptibility really varies substantially among subjects. If some individuals can use their provisions for combating invading agents to throw off effects of a moderate dose of organisms, while weaker subjects will succomb; the ordinary dosage treatment should apply.

The exponential approach obviously simplifies mathematical treatment of data, and in its simpler forms will allow calculation of an LD-50 from only one concentration giving partial mortality. Allowance can be made for varying susceptibility, but in so doing, simplicity if forfeited and advantages over probit analysis seem dubious.

With data of Fort Detrick, Goldberg's approach has given LD-50 estimates very similar to those from probit analysis. The graphic error estimates of his early publications seem inadequate. Where several concentrations give partial mortality, Goldberg's graphic method will yield several LD-50 estimates for the same experiment. These sometimes very incongruously for agents with characteristically low slope.

Critical test comparing the two approaches are very difficult since for ordinary experiments with small numbers results are apt to be quite similar. One possible test involves the form of the untransformed dosage-percentage curve. With the typical probit curve we have an asymetric sigmoid with a weakly defined but real lowe bend. With the exponential we have a single-bend curve of decreasing steepness. Demonstration of a lower bend in the zone of low mortality would be evidence for the probit approach, but would require hundreds of animals. In general, critical test would be expensive and would impede the progress of needed practical tests. We are at present limited largely to gleaning evidence from practical tests.

A preliminary test was afforded in 1953 by Fort Detrick data originated by A. N. Gorelick, of a number of toxic bacterial injections into mice. (See S. B. Job No. 433, Fort Detrick). Some 43 points based on 435 animals were available. If proportion of survival (q) with dosage _n_ is estimated as $(1-a)^n$, then

$$\log q = n \log (1-a)$$

and dosage should be linear in relation to log survival. Significant departure from linearity should suggest that the logical basis of the independent action theory is weak in this material.

On plotting log survival against dose, a gentle curve was suggested by the chart. On fitting, a simple parabola gave a significant gain over a straight line. Statistics are as follows:

Source of Variation	Degree of Freedom	Mean Square
Linear	1	2.44
Quadratic (Additional)	1	0.33
Residual	40	0.03

Another test, not of a definite dosage affect study, but of some assumptions related to those of the independent action theory, was afforded by some of W. C. Patrick's data at Fort Detrick. It involved an encephalomyelitis virus injected intra-cerabrally into mice. A large number of test, made routinely in development work, were available. The agent is very toxic to mice, when injected intracerebrally at 0.03 ml of high dilutions. The regularity of results has led to some thought that any single infective particle reaching the site of action may be fatal.

Following this theory, in the high dioutions allowing survival, the survival is thought to be due simply to the fact that the small sample taken for injection contains no particles. This would imply a Poisson distribution of particles among such samples, with a rather small mean. This would throw us back on the independent action or "one-shot" theory of toxicity.

Patrick's numerous records offered a chance to test this theory. If infective particles have a Poisson distribution among injection samples, and if survival indicates a blank sample, the average number M of units per sample could be estiamted from the proportion of survivors q:

$$q = e^{-m}; \; M = -\ln q.$$

These estimates are quite easily made. Then with two successive concentrations, giving partial mortality; the ratio of two estimates

of m in one test should approximate the dilution ratio (in these cases O. 5 log). This would not be realized exactly in any one comaprison, but with a long series the relation should appear. Filure of the M ratios to agree with the dilution ratios is regarded as evidence against the theory.

For illustration a firaly typical assay of an encephalomyelitis preparation by intra-cerebral injection in mice is taken. Unlike Patrick's series, dilutions were a log apart rather than half a log.

Log dilution	Response	o/o	p	q	Estimated m
7.O	16/16	100.O	1.OOO	O.OOO	--
8.O	16/16	100.O	1.OOO	O.OOO	--
9.O	9/15	60.O	O. 600	O. 400	O. 92
10.O	3/16	18. 8	O.188	O. 812	O. 21

From the first dilution (log is 9.O) showing partial mortality, the value of q is O.O4OO. The theory being tested would say that O. 4 proportion of the injection samples contained zero particles. Solving the equation $q = e^{-m}$ with q taken as O.4O, m comes out as -ln(q) or O. 92. The second dilution similarly treated gives as an estimate of m, O. 21. The ratio is O. 92/O. 21 or about 4. 4. This is far from the dilution ratio of 1O, to be expected if the theory holds.

With the aid of Pvt. Isen a large number of such ratios from Patrick's 1955 and 1956 test were assembled. Logs of computed ratios, from tests where 2 estimates from partial mortality were possible, were assembled and compared with the theoretical O. 5O.

Year	No. Tests Used	Mean Log Ratio Of Estimates of M	95 o/o Confidence Limits
1955	1O4	O. 33	O. 27 - O. 39
1956	166	O. 4O	O. 33 - O. 47

Results do not bear out the theory that a Poisson distribution of infective particles will explain mortality or survival.

To sum up, experience with the independent action model in all-or-none tests at Fort Detrick, has not been very encouraging. Limited tests of the theoretical basis have not sustained the basic theory.

ANALYSIS OF A FUNCTION IN COLLABORATIVE EXPERIMENTATION

Walter D. Foster
Biomathematics Division
Fort Detrick, Frederick, Maryland

I. INTRODUCTION. The usual objective in collaborative or referee experimentation is to make comparisons among the set of participants with the over-all criterion that stations be no more diverse than runs at a station. Thus, station means and variances are the values for interstation comparison. This paper is concerned with the response variable for a particular class of referee experimentation and its analysis.

In this collaborative experiment, each of five laboratories ran a series of aerosol tests in which *P. tularensis* tagged with radioactive phosphorous (P^{32}) was aerosolized in rotating drums and sampled at eight points in time over a 22-hour period. The five laboratories with identical equipment achieved the series of tests at approximately the same time, going to extreme lengths to achieve homogenous methodology. Three treatments were introduced consisting of three relative humidity conditions in the rotating drums of 20%, 50% and 80%. Two aerosols or runs were completed per humidity at each participating laboratory on a randomized basis. It is of interest to note that three separate nations were represented in these five stations.

It was the objective of this experimentation to

(1) Comparestation means,

(2) Compare station variances,

(3) To identify stations whose results did not conform to those of the others,

(4) To examine the station by treatment interaction, i.e., whether the differences between treatments were consistent from one station to another.

II. DEFINITION OF THE RESPONSE VARIABLE. When an aerosol is monitored over a period of time, the measurement usually taken is the concentration at a series of points in time. Thus, the definition of the response variable to be analyzed could be the concentration,

given a particular set of sampling times. However, this concept is likely to ignore the design restriction that only runs are random, not sampling points in a run. A second and better response variable is the function describing concentration and its change in time. Such a function in aerobiology is called a decay function. Previous research has identified a reasonably simple expression which is excellent for summarizing the course of an aerosol in time:

$$C = C_o(t + 1)^{-b} e^{-kt}$$

The usual univariate approach to the analysis of a function such as the one given above would be to analyze separately the parameters of this function, C_o, b, k. However, not only are these parameters known to be correlated because of the design of the experiment but they are also known to be stochastically correlated from one aerosol run to another. Therefore, it is the purpose here to show how the entire decay function, identified as the response variable, can be analyzed and interpreted through the usual analysis of variance technique.

III. <u>ANALYSIS OF VARIANCE OF THE DECAY FUNCTION</u>. With the decay function as the response variable, the following analysis of variance has been accorded this response for the purpose of examining stations levels, variability, and station by treatment interaction. The complete analysis of variance is shown in Table I in detailed form where all of the objectives have been answered. Its construction is given in a separate section.

TABLE I.

A. V. OF DECAY FUNCTION FOR STATIONS AND TREATMENTS

Line	Source	df	MS
15	Mean	3	387.1591
16	Stations	12	.1737
17	A vs Rest	3	.5894
18	Among Rest	9	.0352
19	Treatments	6	.0409
20	S x T	24	.0151
21	Runs in S x T	45	.0195
22	Runs in 20%	15	.0315
23	Runs in 50%	15	.0129
24	Runs in 80%	15	.0118
25	Deviations	150	.0014
26	TOTAL		

The following brief interpretation is accorded the analysis of variance shown in Table I in order to provide specific answers to the objectives of this experiment. Reading from the bottom of the table, the runs have been pooled over stations per treatment affording a test of homogeneity of variance from one treatment to another in lines 22-24. This departure from the original objective is better achieved than the original for estimating station variability because of the limited number of runs per treatment. There is a suggestion that the runs were less homogeneous at the 20% humidity than at the other two. In line 20, it is clear that the station by humidity interaction, if not zero, was small. On the other hand in line 16, differences among stations were obviously large compared to runs in S x T, line 21. The contrast of A versus the remaining stations, line 17, accounted for a large proportion of the station variability, with the variation attributed to the remaining stations being scarcely larger than the variation among trials at a given station. The purpose of this partition in line 17 was to investigate whether the variation among the remaining stations has been reduced to magnitude of trial-to-trial variation. Further partition is in order so long as it could be helpful in identifying and possibly eliminating factors at stations causing station departures.

This brief interpretation was developed completely on the basis of the analysis of variance in Table I. It would be desirable to present a tables of means to accompany the variance analysis. This is the point at which multivariate techniques in general are at a disadvantage, for there is no plainly defined quantity which is easily tabled. Two suggestions are given here as a means by which the interpretation can be visualized; these are first by graphs and secondly by the coefficients of the decay function. The graphs for each station are given in Figure 1 where the values have been averaged over all three humidity conditions. The coefficients computed as estimates of the parameters of the decay function are given below.

Values of Decay Function Constants

	Stations				
	A	B	C	D	E
Log C_o	2.227	2.053	1.879	1.868	1.972
b x 10^1	.868	.994	.053	.406	.839
k x 10^3	.264	.193	.219	.205	.073

It is appreciated that neither of these means of visualizing station differences is perfect; nevertheless, they are suggested here as the best which are easily available.

A few remarks are necessary here before describing in the next section the technique for the analysis of variance of a decay function. The question of auto-correlation always seems to appear in problems in time series such as these. However, it is contended here that because of the function approach the question of auto-correlation of the successive C_i does not arise. Only the residuals are important, and when the decay function is found to provide an excellent summary of the change of concentration in time the residuals may be considered as mutually independent. A second remark has to do with the potential use of the results of this kind of referee experimentation. With the variation noted here and appraised to be acceptable, these data afford a basis for constructing a quality control approach to future aerosol runs in which aberrant points, runs, and even stations may be readily identified.

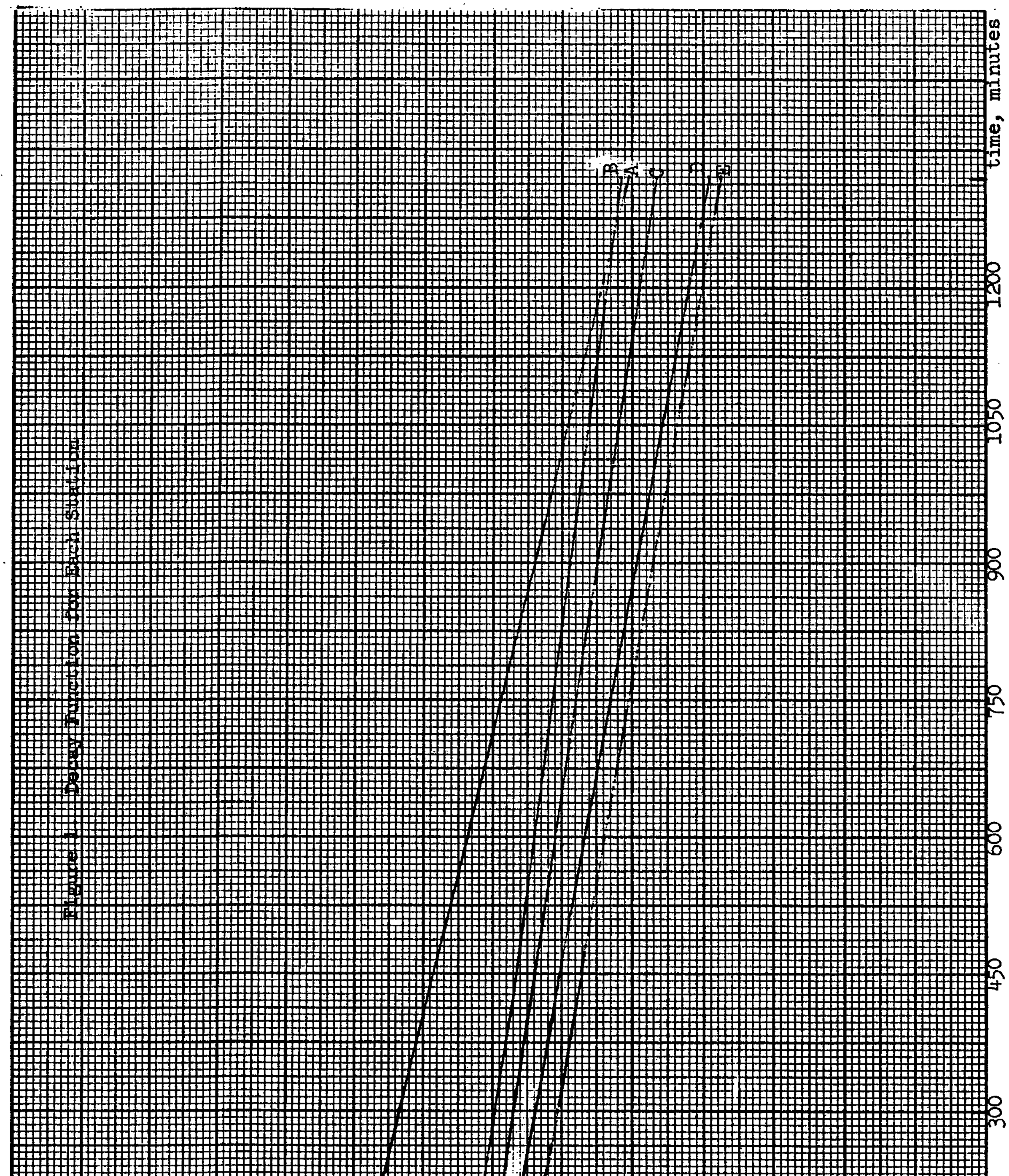


Figure 1. Decay Function for Each Station

IV. CONSTRUCTION OF THE ANALYSIS OF VARIANCE OF A DECAY FUNCTION. On an individual run basis, the familiar partition of variation is obtained as shown in Table II with the exception that no correction is shown separately for the mean -- the p parameters of the decay function are shown together.

TABLE II.

A. V. OF DECAY FUNCTION FOR A SINGLE RUN

Line	Source	df	e. g. df
1	Function	p	3
2	Deviations	n-p	5
3	TOTAL	n	8

The second step is to compute the analysis of variance for each station over the r runs for a given treatment as is shown in Table III. The sum of squares for line 4 are obtained as usual where the function is fitted to the entire set of values for the r runs and the computation is achieved on a per item (or per value) basis. The sum of squares for line 5 is obtained easily merely by summing the sum of squares for line 1 in Table II for the various runs and subtracting line 4. Similarly, line 6, deviations in runs, is obtained by summing the values in line 2 over all runs.

TABLE III.

A. V. OF DECAY FUNCTION FOR A STATION AND A TREATMENT

Line	Source	df	e. g. df
4	Mean	p	3
5	Among runs	p(r-1)	3
6	Deviations in runs	r(n-p)	10
7	TOTAL	rn	16

A small digression may be helpful at this point to explain the degrees of freedom shown thus far in the analysis of variance. The degrees of freedom in line 5 are shown to be the usual degrees of freedom for runs, r-1, multiplied by the number of parameters to be estimated in the decay function. Although these parameters are known not to be independent, they continue to be identified as restrictions in the least squares process for estimation and as such must be deducted as degrees of freedom. It is not likely that a further partition of these degrees of freedom could be achieved in a manner such as to show contrasts among the parameters themselves.

With the introduction of t treatments at a station, the analysis of variance as outlined in Table IV is appropriate for each station, where the partition is basically a nested one. As before, the function is fitted over all points in order to provide the sum of squares due to the function, line 8. The sum of squares for treatments is obtained through a two step procedure. First, the sums of squares shown in line 4 of Table III for each treatment are added. Then the sum of squares for the mean in line 8 is subtracted, the difference being specifically that due to variation among treatments and is entered in line 9. The sum of squares for runs in treatments, line 10, is obtained by summing the sums of squares for each trial separately for that particular treatment, i.e., the sum of lines 5 for that station. They can also be listed in partition as in lines 11 and 12 of Table IV. Similarly, the sum of squares for deviations are obtained by pooling for line 13.

TABLE IV.

A. V. OF DECAY FUNCTION AT STATION A WITH TREATMENTS

Line	Source	df	e.g. df
8	Mean	p	3
9	Treatments	p(t-1)	6
10	Runs in T	pt(r-1)	9
11	in T_1	p(r-1)	3
12	in T_2	p(r-1)	3
	etc	etc	3
13	Deviations	rt(n-p)	30
14	TOTAL	trn	48

The construction of the over-all analysis of variance as shown in Table I continues to be based upon the previous tables in a sort of a building block arrangement. The mean, line 15, is obtained by finding the sum of squares due to the function when fitted to all of the points in the combined collaborative experiment. Line 16 is obtained by a two step procedure: the sum of squares for line 8 in Table IV is summed over the s stations; from this sum of lines 8 the sum of squares in line 15 is subtracted. The difference then represents the sum of squares due to stations averaged over treatments.

The partition of the station sum of squares as initiated in line 17 depends upon which station appears to show the greatest departure from the other stations, following the philosophy given briefly in the interpretation of the example above. Assuming that this identification of the greatest departure can be made from a study of the graphs, line 17 then represents the contrast between the station with the maximum departure and the rest of the stations. This partition is accomplished in a three step procedure as follows. The sums of squares given in line 8 of Table IV are added for the four stations marked as "rest". This sum is entered as line "a" in the ancillary computation table below. The second step is to compute the sum of squares for the function when fitted to all the points represented by the four stations combined as "rest", having excluded the station with the maximum departure from the computation--line "b" below. The third step is to subtract the sum of squares in line "b" from the sum of squares in line "a", giving the "among rest" sum of squares as shown in line "c". Finally, the subtraction of line "c" sum of squares from line 16 is entered in line 17 and is identified as the contrast station A versus "rest". Further orthogonal partitioning for other "departures" can be computed in this fashion.

Ancillary Computation for Table I

Line	Source
a	Sum of line 8 for "rest" stations
b	Mean for "rest"
c	a-b = among "rest" stations

A new computation is required for line 19, the sum of squares due to treatments. This is accomplished by considering all points for the first treatment including those for the various stations and fitting the decay

function. This is achieved for each treatment. These sum of squares are added over the various treatments. From this over-all sum, the value in line 15 is subtracted, giving the variation among treatments averaged over stations.

The interaction term, station by treatment, as shown in line 20, is obtained in the usual way. Briefly, it consists of summing line 4 over all stations and treatments. From this sum are subtracted lines 15, 16 and 19.

Line 21 is obtained easily by summing all lines in Table III. The partition of line 21 as shown in lines 22 and 23 is easily accomplished according to the purpose at hand merely by restricting the summing to the category desired.

Missing values will complicate this analysis and indeed will render the partition non-orthogonal if missing values are not restored to the analysis. Therefore, it is recommended that a simple procedure for estimating these missing values such as computing the value according to the function as estimated from the remainder of the points being inserted with one degree of freedom per missing value being subtracted from the degree of freedom assigned to deviations. Note that in the simpler analyses which are completely nested orthogonality does not depend upon equal numbers.

DESIGN AND ANALYSIS OF ENTOMOLOGICAL FIELD EXPERIMENTS

William A. Brown
Test Design and Analysis Office, Dugway Proving Ground

and

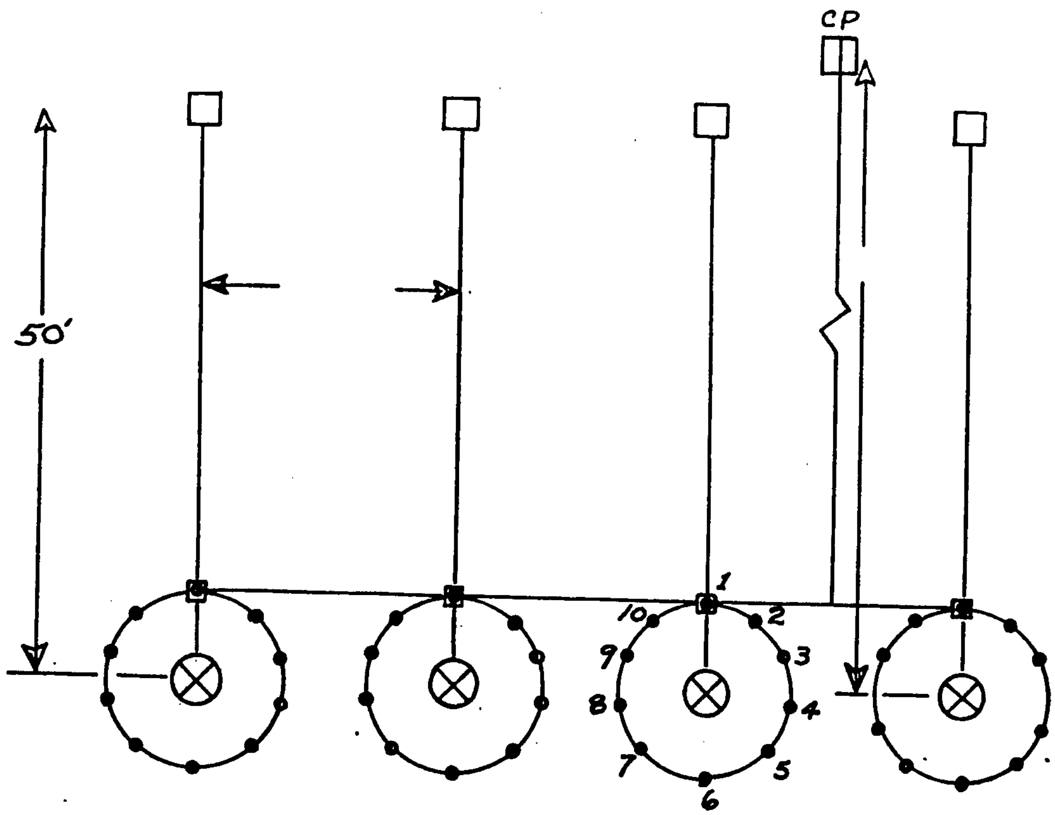
Scott A. Krane
Dugway Field Office, C-E-I-R, INC.

Recently two entomological field experiments were conducted at Dugway Proving Ground. The purpose of the first experiment was to compare the biting propensity of two strains of a species of insect. In each trial, four 15-foot radius circles were scribed, and 10 hosts, randomly selected, were positioned equidistantly along each circumference. The Number 1 position in each circle was oriented to true north. (See Figure 1.) At function time, 100 individuals of the appropriate strain were released at the center of each circle. In two of the circles, the A strain was used; in the other two circles, the B strain. The men were seated on the ground and remained relatively motionless throughout the trial. Sampling consisted of each man recording those bites actually received and entering the total number, for 5-minute intervals, on a data card. Sampling was conducted for 30 minutes following the release unless biting activity continued. In that event, sampling in all circles was extended for additional 5-minute periods until the biting activity had essentially ceased.

Comparisons between strains were thus subject to the variation found among circles. This variation was expected to be appreciably larger than the variation among men on a circle. By the nature of the experimental treatments (strains), however, it was necessary to separate the strains either in space or in time sufficiently that their ranges of biting activity did not overlap. Only in this manner could bites be accurately attributed to one strain or the other. The duplicate circles for each strain represented an effort to partially overcome this inherent insensitivity.

In the analysis of the data, it was considered useful to employ a mathematical model to describe the distribution of the number of bites per host. The simplest model which might conceivably fit the observations is the Poisson, given by:

(1) $$f(x) = e^{-m} m^x/x! \qquad x = 0, 1, \ldots, n,$$



— Local battery field telephone line; □ telephone

○ On-site meteorological sensing station

◘ Boxed recording instruments

╳ Test fixture

● Hosts

Figure 1

where x is the number of bites received by an individual host, $f(x)$ is the probability (or relative frequency) of x bites, and m is an unknown parameter equal to the "long-run" average number of bites per host. If the individuals of a strain are randomly distributed throughout a given area, and if hosts are equally attractive, then the Poisson model should be appropriate.

Previous studies, however, have indicated that the spatial distribution of insects released in this manner is not random (perhaps being influenced by the wind direction, for example), nor are all hosts equally attractive. As a result of these tendencies, the distribution of bites will be "over-dispersed" relative to the Poisson distribution, i. e., the number of hosts receiving a very large number of bites and the number of hosts receiving a very small number of bites will both be larger than the number predicted by the Poisson model, while the number receiving near-average numbers of bites will be smaller.

One of the simplest and most frequently used "over-dispersed" statistical models is the negative binomial, which has the general term:

$$(2) \qquad f(x) = \left(q^{-k}\right)\frac{(k+x-1)!}{x!\,(k-1)!}\,\frac{p^x}{q}, \qquad x = 0, 1, 2, \ldots, n,$$

where q equals $1+p$, and p and k are unknown parameters. Various rationales may be given for the negative binomial.[1] One of the simplest is that the negative binomial is produced by a mixture of Poisson distributions in which the parameter, m, varies according to a "gamma" distribution. While no rationale appears to be particularly compelling in the present problem, the relative simplicity of the negative binomial model and the success with which other investigators have applied it to biological data are taken to justify its use, at least as a working hypothesis.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for over-dispersion with respect to a Poisson distribution by the χ^2 statistic:

---

[1]Bliss, C. F. Fitting the Negative Binomial Distribution to Biological Data. Biometrics, Vol. 9, (2) pp. 176-196.

(3) $$\chi^2_{n-1} = (n-1)s^2 / \bar{x},$$

where n is the sample size (number of hosts, s^2 is the sample variance of the number of bites, and $\bar{x}$ is the sample mean number of bites.[2] The calculated statistic was tested for significance by comparison with the 20 per cent upper tail value of the χ^2 distribution.[3] If the test did not indicate over-dispersion, the data were subsequently fitted to a Poisson distribution and subjected to a χ^2 "goodness-of-fit" test. If the test did indicate over-dispersion, the data were fitted by the method of maximum likelihood[4] to a negative binomial distribution, and then subjected to a χ^2 goodness-of-fit test. All of the above calculations were performed on the IBM 1620 Computer, using a specially prepared FORTRAN program. Fifty-five of 96 sets of 5-minute data showed close agreement with the Poisson distribution. For each of these sets of data, however, the variance was usually larger than the mean, and, consequently, a further comparison with the negative binomial distribution would generally have shown even closer agreement.[5] Therefore, it was decided that, for the purposes of the analysis of variance, an appropriate transformation to stabilize variance for these data would be that derived for the negative binomial:[6,7]

---

[2]Ibid.

[3]For confeni ence of internal calculation on a digital computer, the 20 per cent upper tail χ^2 value was obtained from the approximation:

$$\log_e \frac{\chi^2{n-1}}{n-1} - 1 = -0.038 - 0.452 \log_e(n-1)$$

[4]Fisher, R. A. Notes on the Efficient Fitting of the Negative Binomial. Biometrics, Vol. 9(2), pp. 196-200, 1953.

[5]The Poisson is, in fact, a limiting case of the negative binomial, from which it follows that a negative binomial must fit data at least as well as the Poisson.

[6]Bartlett, M. S. The Use of Transformations. Biometrics, March 1947, Vol. 3(1) pp. 39-52.

[7]Kempthorne, O., Design and Analysis of Experiments. Chapter 8. John Wiley and Sons, Inc., 1952.

(4) $$y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2}).$$

Figure 2, showing a plot of mean versus variance on log-log paper, illustrates the closer agreement with the negative binomial distribution. The diagonal line represents the square root transformation, appropriate for variance stabilization of Poisson distributed data, and the curved line represents the transformation $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ has the value 1.0. (After several guesses of λ, the value of 1.0 was selected since, by eye-fitting, it appeared to reasonably minimize the deviations from the curve. Using the value $\lambda = 1.0$, 47 of the data points lie above the curve, and 49 below.) Subsequent analysis of the data of Experiment 1 using a method of Bliss and Owen[8] for the estimation of a common k, resulted in the estimate

$$k_c = 0.51.$$

Since $\lambda^2 = (1/k)$, the value of λ appropriate for this estimate of k is

$$\lambda = k_c^{-0.5} = 1.4.$$

As shown in Figure 2, the value of $\lambda = 1.0$ obtained graphically agrees reasonably well with that estimated by the method of Bliss and Owen. It can easily be seen in Figure 2 that the data follow more closely to the curved line. However, the dashed line indicates that the logarithmic transformation may be as suitable as the inverse hyperbolic sine. Furthermore, analysis of logarithmically transformed data permits interpretations of results, in terms of ratios of treatment effects, while no such interpretation arises directly from the negative binomial transformation. Therefore, separate analyses of variance were performed, using the two transformations.

An analysis of variance, based on the three-way cross classification of trial, strain, and time period, was performed on each of the following four sets of data:

1. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ equals 1.0 and x is the total number of bites received by a host during a given time period,

[8]Bliss, C. I. and A. R. G. Owen, "Negative Binomial Distributions With a Common K", Biometrika 45, pp. 37-58, 1958.

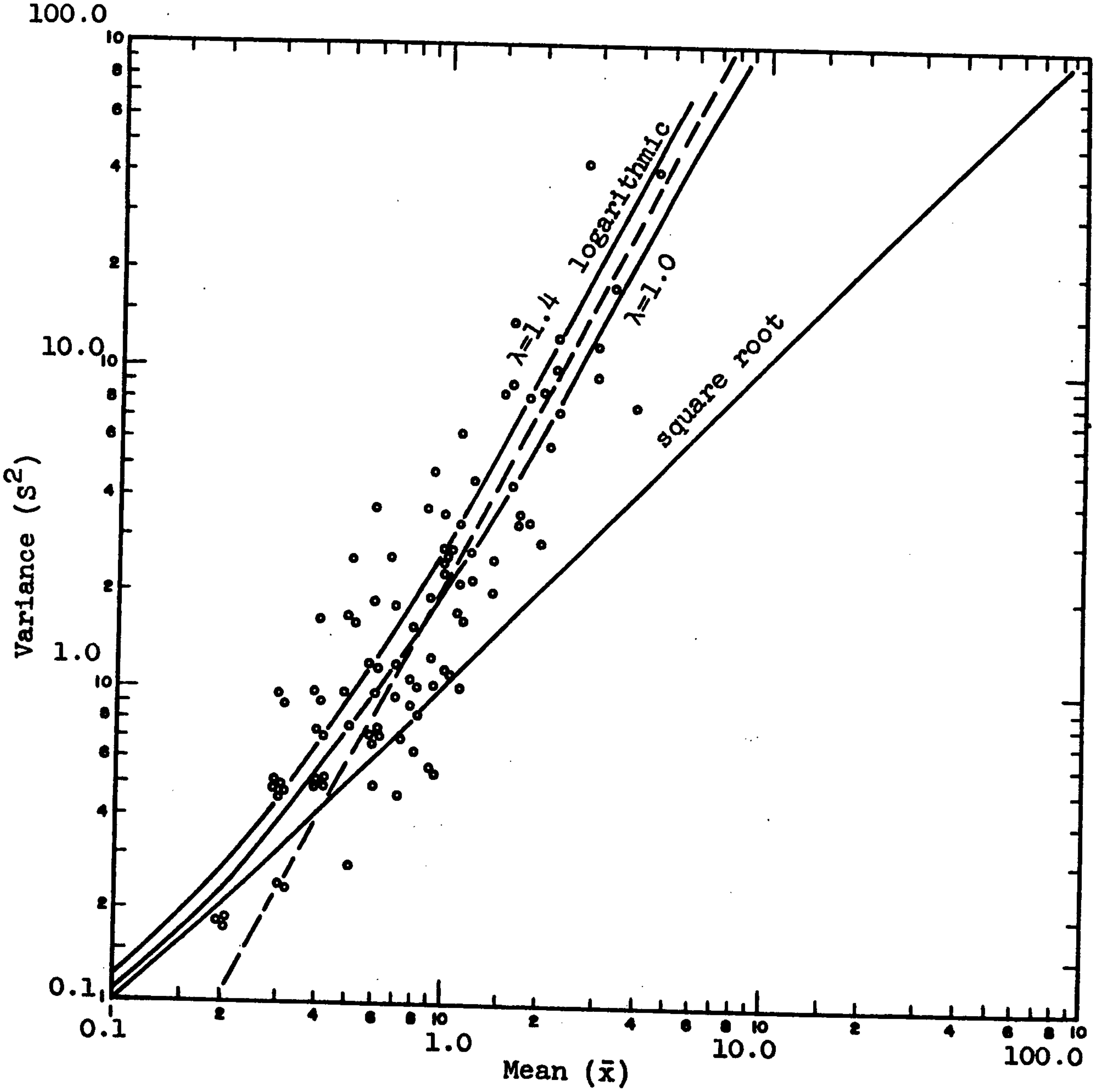

100.0
10.0
1.0
0.1
Variance (s^2)
0.1
1.0
10.0
100.0
Mean ($\bar{x}$)
logarithmic
λ=1.4
λ=1.0
square root

Figure 2

2. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, where λ equals 1.0 and x is the total number of bites received at a circle during a given time period,

3. The values of $y = \log (x + 1)$, where x is the total number of bites received by a host during a given time period, and

4. The values of $y = \log (x + 1)$, where x is the total number of bites received at a circle during a given time period.

The results of each of these analyses are presented in Table 1.

As shown in Table 1, the results obtained in the four analyses of variance were essentially the same. Each analysis indicated that the total numbers of bites obtained during the six times periods were significantly different, and that no significant difference could be detected between strains. In every analysis, however, Error (a) was relatively large, so that the F test, comparing strain effects, was undoubtedly insensitive. As mentioned earlier, the insensitivity of the analyses for strain differences follows unavoidably from the design of these trials, in which strain comparisons could only be made between circles (rather than within circles), and, hence, are subject to the greater variability found from circle to circle as measured by Error (a).

The purpose of the second experiment was to compare the dispersal of two strains of a species of insect as measured by their biting activity.

For this experiment, it was greatly desired that the ambient air temperature and windspeed range of an A-B strain pair of trials be as similar as possible. However, because of the small number of men available concurrent testing of the two strains could not be accomplished. Therefore, whenever possible, two trials were conducted each day--one trial using the A strain and the other, following as soon after as practicable, employing the B strain.

In each trial, four concentric circles were used, designated Circles A, B, C, and D with radii equal to 100, 200, 300 and 400 feet. Eight men were positioned equidistantly around each circumference of Circles A, B, and D, and 16 men were positioned equidistantly around the circumference of Circle C. (See Figure 3.) At function time, 1000 individuals of the appropriate strain were released at the center of the concentric configuration, and the men, seated and facing the release point, recorded

Table 1

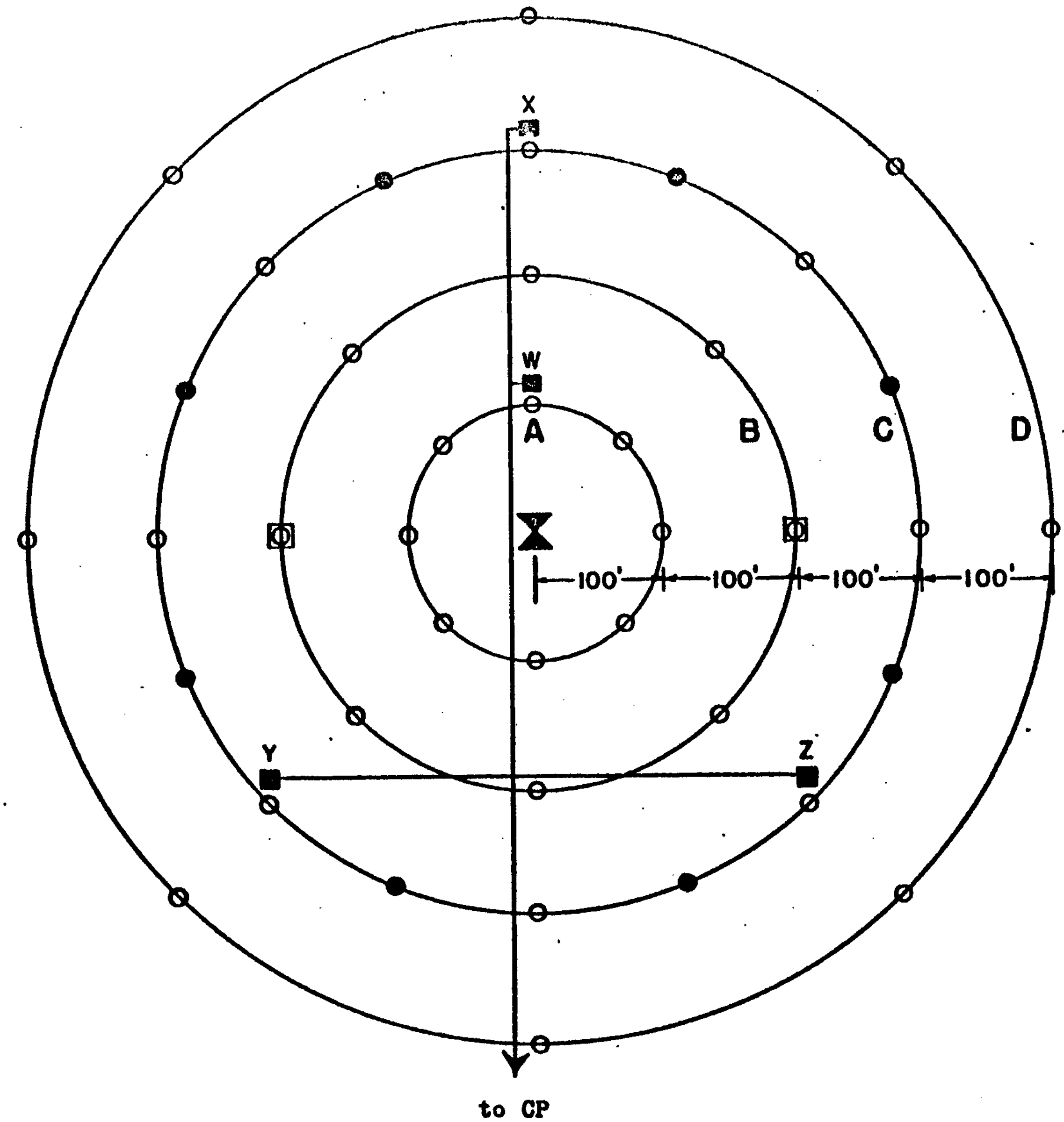
SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMATION*							
		1		2		3		4	
		Mean Square	F Value	Mean Square	F Value	Mean Square	F Value	Mean Square	F Value
Trial, T	5	0.076396		0.054999		0.279571		0.216414	
Strain, S	1	0.240379	3.27	0.240754	2.39	0.894409	3.30	0.944946	2.48
T x S	5	0.017313	0.236	0.25430	0.253	0.065268	0.241	0.095473	0.251
Error (a)	12	0.073442		0.100546		0.270806		0.380320	
Time period, P	5	0.571952	22.8**	0.716051	51.8**	2.073743	60.4**	2.743388	54.1**
P x T	25	0.028863	1.15	0.026287	1.90	0.105622	3.08**	0.101577	2.00*3
P x S	5	0.011180	0.445	0.007883	0.570	0.041707	1.21	0.027349	0.540
P x T x S	25	0.016094	0.641	0.019936	1.44	0.058558	1.71	0.075415	1.49
Error (b)	60	0.025108		0.013830		0.034306		0.050667	
Sub-total	143								
Hosts/Circles, C	216					0.147148			
Hosts/P x C	1080					0.029697			
Total	1439								

Trials A-3, A-8, and A-9 were omitted from this analysis because of insufficient data.

*1 denotes $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$ transformation of 5-minute total bites received by each host; 2 denotes $\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$ transformation of 5-minute total bites received at each circle; 3 denotes log $(x + 1)$ transformation of 5-minute total bites received by each host; and 4 denotes log $(x + 1)$ transformation of 5-minute total bites received at each circle.

**Significant at the 1.0 per cent level.

*3 Significant at the 5.0 per cent level.



■ Field telephone

— Field telephone line

O Sampling station

● Augmenting sampler position

□ 2-meter wind speed and direction station

Release point

Figure 3

biting activity for at least 30 minutes. The sampling proceedures were the same as those used for the first experiment.

For each 5-minute time period, the mean and variance of the reported bites at each circle were estimated. Each set of data was then tested for "over-dispersion" with respect to a Poisson distribution in the same manner as in the first experiment.

The results indicated that there was nearly always a departure from the Poisson distribution, in the direction of higher variance and "over-dispersion." In addition, 45 of the 70 sets of 5-minute data showed agreement with the negative binomial distribution at a nomial 95 per cent confidence level. Further, from an examination of the plot of the mean versus the variance (see Figure 4), it did not appear that the data would fit any other distribution more consistently. Therefore, it was decided that, for the purposes of the analysis of variance, a suitable transformation to normalize these data would be:

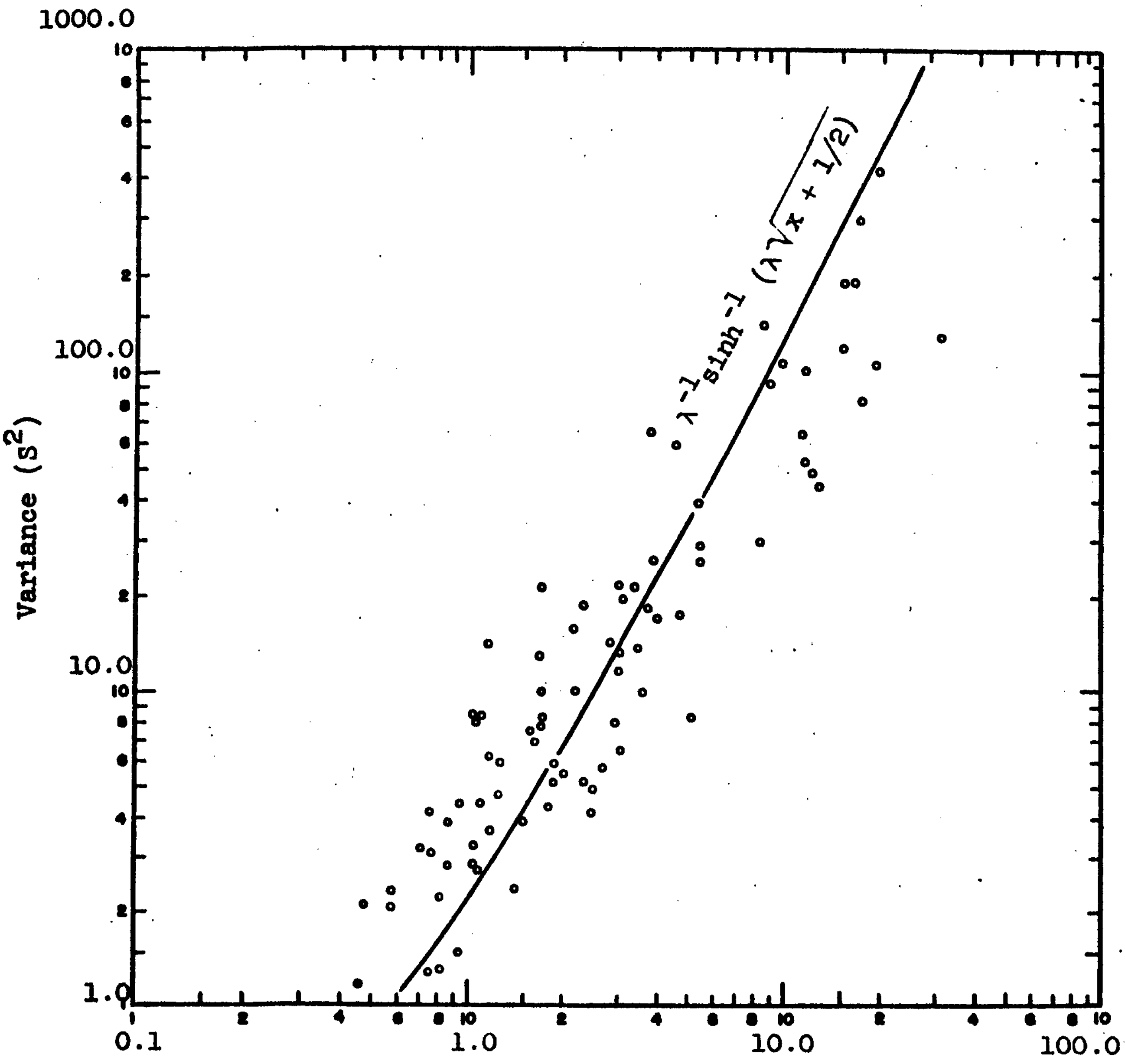
$$y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2}).$$

After several guesses of λ, and, subsequently, fitting the data by eye to $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$, it appeared that a reasonable estimate that would minimize the deviations from the curve was $\lambda = 1.0$. Using this value, 42 of the data points lie above the curve, and 43 below.

An analysis of variance, based on the four-way cross classification of day, strain, circle, and time period, was performed on each of the following sets of data:

1. The values of $y = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ where $\lambda = 1.0$, and x is the total number of bites received during a given time period at Circles A, B, and D, and one-half the total received at Circle C.

The results of these analyses are presented in Table 2.



1000.0
100.0
10.0
1.0
Variance (s^2)
$\lambda^{-1}\sinh^{-1}(\lambda\sqrt{x + 1/2})$
0.1
1.0
10.0
100.0
Mean ($\bar{x}$)

Figure 4

Table 2: Results of the Analyses of Variance of Transformed Bite Data

SOURCE OF VARIATION	DEGREES OF FREEDOM	RESULTS OF ANALYSIS OF VARIANCE FOR INDICATED TRANSFORMED DATA*			
		1		2	
		Mean Square	F Value	Mean Square	F Value
Day, D	1	0.342065		0.331695	
Strain, S	1	0.265038		0.241103	
Error (a)	1	0.271097		0.225376	
Circle, C	3	1.020387	50.5 **	1.199593	71.9 **
C x S	3	0.044502	2.20	0.044751	2.68
Time Period, T	9	0.670518	33.2 **	0.623410	37.4 **
T x S	9	0.115344	5.70**	0.108279	6.49**
T x C	27	0.044332	2.19**	0.043864	2.63**
T x C x S	27	0.011870	0.587	0.014230	0.853
Error (b)	78	0.020225		0.016677	
Total	159				

*1 denotes the data resulting from the $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ transformation of total bites received at a circle during a given time period; and 2 denotes the data resulting from the $\lambda^{-1} \sinh^{-1} (\lambda\sqrt{x + 1/2})$ transformation of total bites received during a given time period for Circles A, B, and D, and one-half the total bites received at Circle C.

** Significant at the 1.0 per cent level.

As shown by the F values in Table 2, the second analysis, adjusting for the augmented sampling on Circle C, was the more sensitive. Both analyses, however, showed circle, time period, T x S, and T x C to be highly significant.

Using the second set of transformed data, a further investigation of circle, T x S, and T x C was made in the following way. From the analysis of the transformed values:

(5) $$y_{ijk} = \lambda^{-1} \sinh^{-1} (\lambda\sqrt{x_{ijk} + 1/2}) , \quad \lambda = 1.0$$

$$= \sinh^{-1}\sqrt{x_{ijk} + 1/2},$$

where x_{ijk} equals the total number of bites received by the i-th host at the j-th circle during the k-th time period. The mean values, $\bar{y}_{jk}$, of the transformed variables were obtained. These mean values are related to the estimated true average number of bites received at the j-th circle during the k-th time period, $\hat{m}_{jk}$, by

(6) $$y_{jk} = \sinh^{-1}\sqrt{\hat{m}_{jk} + 1/2}, \text{ hence}$$

$$\hat{m}_{jk} = (\sinh \bar{y}_{jk})^2 - 1/2.$$

Relationships between $\bar{y}_{jk}$ and circle radius R_j, were sought. The best simple relationship found was:

(7) $$\bar{y}_{jk} = a - b \log_e R_j,$$

where a and b are regression constants determined by the method of least squares.

(8) Then,
$$\hat{m}_{jk} = \left[\sinh (a - b \log_e R_j)\right]^2 - 1/2,$$

(9) $$= \left[1/2(e^{a-b \log_e R_j} - e^{-a+b \log_e R_j})\right]^2 - 1/2,$$

(10) $$= \left[1/2(e^a e^{-b \log_e R_j} - e^{-a} e^{b \log_e R_j})\right]^2 - 1/2$$

$$(11) \qquad = \left[1/2\left(e^{a}R_j^{-b} - e^{-a}R_j^{b}\right)\right]^2 - 1/2,$$

which can be approximated by:

$$(12) \qquad \hat{m}_{jk} = \frac{e^{2a}}{4R_j^{2b}} - 1/2,$$

since $e^{-a}\left(R_j^{b}\right)$ is small relative to $e^{a}\left(R_j^{-b}\right)$.

For each 5-minute time period, the transformed data were summed with respect to circle and strain. These values were then fitted to the above regression model, and the average values of the various a's and b's determined. Subsequently, for each strain, the true average number of bites was estimated for each circle during the various time periods. These latter values are presented in Table 3.

Table 3: Estimated True Average Number of Bites of A and B Strain at the Various Circles During Given Time Periods.

STRAIN	TIME PERIOD (Minutes)	ESTIMATED TRUE AVERAGE NUMBER OF BITES AT INDICATED CIRCLE DURING GIVEN TIME PERIOD			
		Circle A (100 feet)	Circle B (200 feet)	Circle C (300 feet)	Circle D (400 feet)
	0 - 5	123	15	4	1
	5 - 10	145	26	10	5
	10 - 15	119	28	12	7
	15 - 20	120	29	12	7
A	20 - 25	82	20	8	4
	25 - 30	39	15	8	5
	30 - 35	22	10	6	4
	35 - 40	16	8	5	4
	40 - 45	8	6	5	4
	45 - 50	6	4	4	3
	0 - 5	99	25	11	6
	5 - 10	132	35	16	9
	10 - 15	99	32	16	10
	15 - 20	59	21	12	8
B	20 - 25	39	19	13	10
	25 - 30	19	9	5	4
	30 - 35	9	6	5	4
	35 - 40	2	2	2	2
	40 - 45	4	2	1	1
	45 - 50	0	0	0	0

As shown in Table 3, the expected number of A strain bites at Circle A during each of the various time periods is greater than that for B strain; however, the difference, in general, is not appreciable. At Circles B, C, and D, there appears to be no important difference between the number of bites. It was, therefore, concluded that the spatial dispersion of the two strains was comparable, as indicated by the nonsignificance of the C x S interaction.

The significance of the T x S interaction indicates a difference between strains with respect to the temporal dispersion characteristics. Generally speaking, the biting activity of strain B appeared to exhibit a more pronounced peak in time and a slightly earlier decline.

Unfortunately, no satisfactory model for the characterization of the biting activity as a function of time has been found by the authors. It is hoped that such a model may yet be developed.

COMPARISON OF APPROACHES TO OBTAINING A TRANSFORMATION MATRIX EFFECTING A FIT TO A FACTOR SOLUTION OBTAINED IN A DIFFERENT SAMPLE

Cecil D. Johnson
U. S. Army Personnel Research Office

BACKGROUND. The importance of physical proficiency measures to the selection and evaluation of Army personnel can scarcely be questioned. Examination of the duty assignments prevalent in various Army jobs indicates clearly that physical strength, endurance and coordination are often highly important factors in job success. At the United States Military Academy in particular, considerable attention has been devoted to physical training and to the measurement of various aspects of physical abilities or physical proficiency among cadets at West Point. Various tests of physical proficiency were introduced in the physical aptitude entrance examination procedure or studied for possible use.. They have been examined both as individual measures and as component parts of various batteries. Several factor analyses of large batteries of physical proficiency measures, physical education, grades, and other variables were accomplished in previous studies. These studies had as their objective the identification of basic underlying physical ability variables that possess the simplifying statistical characteristics frequently referred to as simple structure. These basic variables, or factors, aid in understanding the nature of the scores, in eliminating duplicating measures, and in suggesting new tests.

The several factor solutions available for comparison contain numerous variables in common, other similar variables (as when a 25-yard dash is substituted for a 30-yard dash), and still other variables which are unique for a particular solution. This paper considers several methods for comparing solutions obtained in these separate studies involving physical proficiency and related measures.

The problem of approximating in a second sample, a rotated factor solution originally obtained in a previously analyzed sample is also present in another Army Personnel Research Office research study currently in the final computing phase. This study involved thirty-one psychological tests. Some of these tests are measures of intellectual ability, others are measures of cognitive information, and others are non-cognitive measures in the "personality" domain. As is typical with Personnel Research Office factor analysis studies, the objective was the identification of constructs which would predict the performance of soldiers on the job. In this case the job was that of an enlisted Infantryman and the measure of performance was obtained from ratings by superiors and peers at the close of maneuvers in

Germany. The tests had been administered to the enlisted men on their entry into the Army.

An initial principal component factor solution was transformed by an orthogonal matrix so as to provide simple structure. The initial factor solution can be described as a matrix whose elements are the correlations between the tests and standard length orthogonal reference vectors. This solution usually provides parsimony in that a relatively small number of reference vectors is needed to closely approximate the test correlation matrix when the factor matrix is post-multiplied by its transpose. However, the psychologist wants the reference vectors, or factors, to have additional properties implied by the concept of a simple structure. If simple structure is present among the reference vectors, each reference vector has high correlations with a few tests and approximately zero correlations with the remainder. Furthermore, the tests with which a particular reference factor has a high relationship will be relatively independent of the other reference vectors. It is apparent that the presence of simple structure permits the psychologist to interpret the reference vectors in terms of his test, and if the orthogonality of the reference vectors is retained, as when the transformation matrix is orthogonal, all the original parsimony of the initial principal component factor solution is retained. Psychologists usually refer to the process of transforming a solution to simple structure as rotation, and call the transformed solution a "rotated" solution.

In the factor analysis of psychological tests described above, the rotated solution, when extended to the rating variables, displayed a very interesting relationship between the rotated factors and the performance measures. One cognitive factor and one non-cognitive factor predicted performance while all other factors had a zero relationship with performance. It became a matter of considerable interest to determine whether these relationships could be verified in an independent sample where properties of the sample had not been used to determine the particular transformation used to obtain the rotated factor solution. Both factors retained their validity in the cross (independent) sample, but an additional factor (previously non-valid) also displayed a smaller amount of validity. In this study the factor validities in the first sample were fairly well replicated in the second sample.

Thus, both studies, the one involving physical proficiency variables and the one involving psychological tests, require an initial factor solution in a second independent sample, the transformation of this solution to one

approximating the rotated solution in the first sample, and finally, the extension* of the transformed solution to non-overlapping predictor variables and/or criterion variables. The criterion variables may well overlap across the two studies but should be withheld from the initial factor analysis for two reasons:

(1) It is desirable that the factors be entirely defined by predictor variables.

(2) The validity of the transformed factors are being determined in the independent, or cross sample. Thus, the definition of the factors in the cross sample must be independent of the criterion variables.

B. F. Green has reported a method for computing an orthogonal transformation matrix which will minimize the sum of squares of the differences between the transformed matrix and the matrix to be fitted. However, his derivation does not generalize so as to provide an orthonormal transformation that can utilize more reference vectors in the cross sample than are in the matrix to be fitted.

If the investigator is confident that the initial cross sample factor solution does not have a rank which exceeds the rank of the solution to be fitted, this orthogonal transformation is clearly suitable. On the other hand, if in the cross sample there is likely to be considerable variance common to two or three variables that is not explained by the more general common factors utilized in the initial sample, the advantages of an orthonormal solution become apparent.

APPROACH AND RESULTS. Thus, in obtaining the transformation matrix necessary for fitting K factors, the investigator has a choice of using a method which obtains the best orthogonal transformation matrix applicable to the first K factors, or he can choose to use a non-square orthonormal transformation matrix which can be applied to a full factorization, i.e., to as many factors as there are variables. The first method, using an orthogonal transformation, requires the fitted solution to reproduce

---

*This factor extension is accomplished by post multiplying the m x n matrix of correlation coefficients (between the overlapping and non-overlapping) by $A_y D_y^{-1/2}$, where R_y is the matrix of correlation coefficients among the n overlapping variables and $A'_y R_y A_y = D_y$; $A'_y = A_y^{-1}$; D_y = eigen values.

the cross sample correlation matrix to the full extent possible with a principal component solution. The second method, using an orthonormal transformation, provides a less exact reproduction of the correlation matrix, but permits a better fit to the reference factors -- if the dimensionality of the experimental variable space exceeds the number of factors.

Since in the physical proficiency study the number of common variables, for some of the comparisons, was large as compared to the number of factors being fitted, and the initial solutions had been obtained using a correlation matrix involving even more variables, the non-square, orthonormal transformation was utilized. A description of this technique is provided in the hand-out.

The three rotated factor solutions to which the initial solutions in the cross-samples were fitted, tended to have smaller communalities than the four cross-sample fitted solutions. This is possibly explained by two things. The methods of obtaining the initial solutions (that were subsequently rotated in the reference samples) were less efficient that the principle component method used for the initial cross sample solutions. Also, the initial factor solutions in the first sample were obtained to span the non-common variables as well, whereas the cross sample solutions were obtained on the common variables only.

In comparing the use of the orthonormal as compared to an orthogonal transformation, it becomes a trade off between the better fit to the inter-correlation matrix obtained by using an orthogonal transformation and the better fit to the rotated factor solution possible under certain circumstances with the orthonormal transformation. The differences between the two methods in regard to fitting the rotated factor tend to diminish as the number of factors involved increases. On the other hand, the advantage possessed by the orthogonally transformed solution in reproducing the inter-correlation matrix increases. Thus it is clear that the value of the orthonormal transformation as compared to the orthogonal transformation is least likely when the number of factors in the initial rotated solution is large. However, the number and nature of the non-overlapping variables in the two studies is also important.

Since the extension of the transformed solution to the non-common variables is an important aspect of these studies, the reproduction of the intercorrelations between the extended factors and these additional variables is an important consideration.

It is interesting that the advantage of the orthogonally transformed solution for reproducing the intercorrelation matrix did not always hold among the non-overlapping variables. This underlines the fact that the advantage of the orthogonal transformation matrix in reproducing the correlation matrix in the cross sample is partly due to its more efficient capitalization on sampling error. This sampling error effect is further underlined by the fact that for the orthonormal transformed solution, for one sample, the elements of the residual matrix involving the non-common variables were smaller than the elements of the corresponding matrix involving the common variables. This was, of course, just the opposite of the results obtained from the orthogonally transformed factor solution. However, while the initial unrotated solution extended to the non-common variables, necessarily possesses the maximization properties of the initial principal component solution for only the common variables, the advantage, while reduced, was still present for non-common variables in the larger samples.

The two following questions were raised at the conclusion of the two USAPRO presentations:

(1) Have factors (i. e., factor pure tests) proved to be good predictors of Army performance criteria?

(2) What is the advantage, for prediction, of using orthogonal predictors over the original correlated predictors if optimal weights are applied?

The two questions are closely related in that they are both concerned with the immediate application of factor analysis results to the practical problem of predicting personnel performance. USAPRO has considerable research evidence indicating that tests developed to measure factors do not predict performance as well as factorially complex tests developed to predict a specific Army performance measure. We have very little evidence bearing on our own factor measures, since, on theoretical grounds, we have not expected factor measures to have immediate use as predictors. Factors are useful constructs because of their simplified (i. e., more easily understood) relationships with psychological or physiological measures and the other factors. Thus, factor scores are useful in experimentally testing hypotheses relating carefully defined psychological content of a measure to human performance, and the factor concept has general usefulness for the better understanding of the psychological content of a battery of tests. It is not expected that factors will have immediate usefulness as operational predictors.

APPENDIX I

Formulae and Notation

I. Certain letters will be used consistently to denote specific kinds of matrices. Different matrices of the same type will be discriminated by their subscripts.

- R — a gramian matrix whose elements are product moment correlation coefficients.
- P — a principal component factor solution.
- A — an orthogonal eigen vector matrix derived from a gramian matrix.
- D — eigen value matrix.
- F — a factor solution other than a principal component factor solution.
- T — a transformation matrix whose elements are cosines of the angles between reference vectors (factors).

II. The following formulae relate several of the above matrices:

$$A'RA = D, \quad AD^{1/2} = P, \quad PP' = R, \quad P'P = D$$

$$FF' = R, \quad F'F \neq D$$

$$F_Y T_1 = F_X ; \quad T_1 = (F'_y F_y)^{-1} F'_y F_x$$

III. Table 1

A Sectioned Matrix Whose Elements are Projections*
Involving the Row and Column Variables

(As computed in the second sample)

		Rotated Factors from 1st Sample	PC Factors in Space Defined by Rotated Factors from 1st Sample	Experimental Variables
		$F_{x1} \cdots\cdots F_{xk}$	$F_{o1} \cdots\cdots F_{ok}$	$Y_1 \cdots\cdots Y_n$
Rotated Factors From First Sample $(k \leq n)$	F_{x1} ⋮ F_{xk}	R_L $R_L = T'_2 T_2$ $P_y T_2 = F_x$ $T_2 = D_y^{-\frac{1}{2}} A'$ **	P_{XL} $P_{xL} = A_L D_L^{\frac{1}{2}}$ $P_{xL} = R_L A_L D_L^{-\frac{1}{2}}$ $A'_L R_L A_L = D_L$	F'_x
Principal Component Factors (PC Factorization of R_y)	P_{y1} ⋮ P_{yn}	T_2 $T_2 = D_y^{-\frac{1}{2}} A'_y F_x$ $T_2 = P_y^{-1} F_x$	T_3 $T_3 = T_2 A_L D_L^{-\frac{1}{2}}$	P_Y $P'_y = D_y^{\frac{1}{2}} A'_y$ $A'_y R_y A_y = D_y$
Experimental Variables	Y_1 ⋮ Y'_n	F_X $F_x = F_y T_2$ $F_x = P_x A'_x$ $A_x(F'_x F_x)A_x = D_x$	F_{YL} $F_{yL} = F_y T_3$ $F_{yL} = F_x A_L D_L^{-\frac{1}{2}}$	R_Y

* These projections are cosines when both row and column vectors are of unit length.

** If $k = n$, $R_L = F'_x R_y^{-1} F_x$. However, in almost all practical situations, k, the number of rotated factors, will be considerably smaller than n, the number of variables.

IV. Green's Procedure

Problem: In order to fit a factor solution of k factors (i.e., P_{yk}) in the second sample to a rotated factor solution, F_x, in the first sample (i.e., $P_{yk}T_o = F_x$), compute T_o such that tr $(P_{yk}T_o - F_x)'(P_{yk}T_o - F_x)$ is minimized, while meeting the side constraint that $T_o' = T_o^{-1}$.

Solution:

(a)* $T_o = (P_{yk}' F_x F_x' P_{yk})^{-1/2} P_{yk}' F_x$, P_{yk} contains the k columns of P_y, the complete principal component factorization of R_y, corresponding to the k larger roots of D_y. $P_y = A_y D_y^{1/2}$ where $A_y' R_y A_y = D_y$ and R_y is the product moment correlation matrix for the experimental variables in the second sample.

(b) T_o can also be computed from the least square transformation, $T_1 = D_{yk}^{-1} P_{yk}' F_x$

$$T_o = (D_y T_1 T_1' D_y)^{-1/2} D_y T_1$$

(c) A slightly different orthogonal transformation matrix, T_p, can be derived by directly orthogonalizing the T_1 as follows:

$$T_1'X = T_p' \quad , \quad T_p'X = T_1^{-1}$$

$$T_1'X = T_1^{-1} X^{-1} \quad , \quad X = (T_1 T_1')^{-1/2}$$

*Green, B. F. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, 1952, 17, 429-440.

$$T_p = (T_1 T_1')^{-1/2} T_1$$

Properties of T_o:

The orthogonally transformed factor solution, $P_{yk} T_o$, retains the maximum reproduction of R_y (i.e., $P_{yk} T_o T_o' P_{yk}' = P_{yk} P_{yk}'$). When communalities are substituted for ones in the diagonals of R_y and if the rank of R_y becomes K, this is undoubtedly the best procedure for fitting F_x.

V. Alternative Procedure

A non-square (orthonormal) transformation matrix permitting the full utilization of P_y can be developed as follows:

a. Whereas T_1 in the previous model provided a least square fit, T_2 in the model below provides an exact fit (since P_y has an inverse while P_{yk} does not when $k < n$.)

$$P_y T_2 = F_x ; \quad T_2 = P_y^{-1} \cdot F_x = D_y^{-1/2} A_y' F_x$$

b. The factor matrix F_x, computed and rotated in Sample 1, cannot usually be obtained by an orthogonal or orthonormal transformation of P_y, computed in Sample 2. The transformation of $P_y T_2$ into a solution within an orthogonal frame F_{yL}, can be accomplished as follows:

$$F_{yL} = P_y T_2 \; (A_L D_L^{-1/2})$$

c. While the matrix F_{yL} contains K column vectors (factors) spanning the same space as the oblique factors in F_x, the orthogonali-

zation was not accomplished in such a way as to maximize the fit of F_{yL} to F_x. The additional transformation required to effect this fit can be accomplished by using Green's procedure to obtain orthogonal matrix T_{o2} as follows:

$$F_{yL}\, T_{o2} = F_x$$

$$T_{o2} = (F'_{yL} F_x F'_x F_{yL})^{-1/2} F'_{yL}\, F_x$$

$$T_{o2} = \left[D_L^{-1/2}\, A'_L\, (F'_x F_x)^2\, A_L\, D_L^{-1/2}\right]^{-1/2}\, D_L^{-1/2} A'_L (F'_x F_x)$$

d. Thus the ortnonormal transformation matrix T_m which minimizes the trace $(P_y\, T_m - F_x)'(P_y T_m - F_x)$, where T_m is an $n \times k$ orthonormal matrix is,

$$T_m = T_2 A_L D_L^{-1/2} \left[D_L^{-1/2}\, A'_L\, (F'_x F_x)^2\, A_L D_L^{-1/2}\right]^{-1/2}\, D_L^{-1/2}\, A'_L (F'_x F_x)$$

VI. An orthonormal transformation matrix T_n providing a least square fit of T_n to T_2, as compared to the fit of F_{yL} to F_x in Part V, can be provided as follows:

a. The conversion of T_2 to an orthonormal matrix, T_3, spanning exactly the same space can be accomplished as follows:

$$T'_2 T_2 = R_L$$

$$T_3 = T_2 A_L D_L^{-1/2} \quad ; \quad T'_3 T_3 = I$$

b. While T_3 is an orthonormal matrix, it is not the orthonormal matrix with the best least square fit to T_2. There is still the need to minimize the trace of $(T_3T_{o2} - T_2)'(T_3T_{o2} - T_2)$, where T_{o2} is an orthogonal matrix. This can be accomplished by making use of Green's procedure described under IV(a).

$$T_{o2} = (T_3'T_2T_2'T_3)^{-1/2}T_3'T_2$$

$$T'_3T_2 = P_{xL}^{-1}T_2'T_2 = P_{xL}^{-1}R_L = P'_{xL} \ ; \quad \text{since } P_{xL}^{-1} = D_L^{-1/2}A_L'$$

$$T_{o2} = (P'_{xL}\ P_{xL})^{-1/2}P'_{xL} = D_L^{-1/2}\ P'_{xL}$$

$$T_{o2} = A_L'$$

c. Thus the $n \times k$ orthonormal transformation matrix T_n which minimizes (in the least square sense) trace $(T_n - T_2)'$ $(T_n - T_2)$ is equal to,

$$T_3T_{o2} = T_2A_LD_L^{-1/2}A_L' = T_2\,(T'_2T_2)^{-1/2}$$

$$T_n = T_2\,(T'_2T_2)^{-1/2}$$

Note the similarity in the form of the computing formulae used to describe T_n and T_p. $T_p = (T_1T_1')^{-1/2}\ T_1$.

APPENDIX II

The Comparison of Transformation Matrices

I. Method

The variables included in the rotated factor solution in the first sample are designated by x and the rotated factor solution by F_x. The same (i.e., overlapping) variables in the second sample are designated by y and the non-overlapping variables by z. The factorization of R_y and R_z are accomplished as $P_y = R_y A_y D_y^{-1/2}$ and $F_z = R_{yz} A_y D_y^{-1/2}$. The transformed factor matrices in Sample 2 are $F_{yr} = P_y T$ and $F_{zr} = F_z T$. Each transformation matrix, T, computed by the methods described in Appendix I, is evaluated by determining the fit of F_{yr} to F_x and the reproduction of R_y by $F_{yr}F'_{yr}$, R_{yz} by $F_{zr}F'_{yr}$ and R_z by $F_{zr}F'_{zr}$. This is determined by comparing (for the different T matrices) the traces of the following product matrices:

$$(F_{yr} - F_x)'(F_{yr} - F_x) \;,\; (F_{zr}F'_{yr} - R_{yz})'(F_{zr}F'_{yr} - R_{yz})$$

$(F_{zr}F'_{zr} - R_z)'(F_{zr}F'_{zr} - R_z)$, and, after setting diagonal elements of $F_{yr}F'_{yr}$ and R_y equal to zero, $(F_{yr}F'_{yr} - R_y)'(F_{yr}F'_{yr} - R_y)$.

II. Results

The sums of squares of the residual matrices, computed as the traces of the matrices indicated in Part I above, are provided in Table 1 for a study involving physical proficiency measures. The x sample consisted of 254 West Point Cadets of the class of 1949. The y sample contained 294 West Point Cadets of the class of 1964. Table 2 relates to a study involving the following x and y variables (in samples 1 and 2 respectively): 15 "Personality" tests, 9 information tests, and 8 mental aptitude tests. Five rating variables based on performance as Infantryman, make up the z variables. Sample one (x variables) had 550 examinees and sample two(y variables) had 375 examinees.

The rank ordering of the magnitudes for the various entries in Tables 1 and 2 can be readily predicted from the algebraic formulations of the T's. The relatively efficiency for fitting F_x, going from high to low, is T_m, T_n, T_o, T_p. The relative efficiency for reproducing the R matrices is the same for all T's which are either orthogonal or capable of being linked by an orthogonal transformation. Thus all the orthogonal T matrices have more efficiency for reproducing R_y, when applied to PC solutions of R_y, than do the orthonormal transformations.

Table 1

Comparison of Transformation Matrices Computed on a Sample of 294 Examinees

(Physical Proficiency Variables)

Residual Matrices	Number of Elements Contributing to the Sums of Squares Reported as Entries in this Table	Total Sums of Squares of Elements in Residual Matrices						
		ORTHOGONAL Transformations				ORTHONORMAL Transformations		
		Communalities in Diagonals of R_y		Ones in Diagonals of R_y		Ones in Diagonals of R_y		Communalities in Diagonals of R_y
		T_o	T_p	T_o	T_p	T_m	T_n	T_m
$(F_{yr} - F_x)$	19 x 9 = 171	2.1381	3.1934	5.2665	5.7615	1.0943	1.5233	1.6758
$(F_{yr}F'_{yr} - R_y)$, other than diagonal elements	19 x 18 = 342	.2268	.2268	.8496	.8496	5.1851	5.1851	1.2382
$(F_{zr}F'_{yr} - R_{zy})$	19 x 11 = 209	.2953	.2953	.4129	.4129	3.0342	3.0342	.8050
$(F_{zr}F'_{zr} - R_z)$	$(11)^2$ = 121	2.1925	2.1925	3.6127	3.6127	6.2471	6.2471	2.3205

Table 2

Comparison of Transformation Matrices Computed on a Sample of 375 Examinees

(Personality, Mental Aptitude, Information, and Rating Variables)

Residual Matrices	Number of Elements Contributing to the Sums of Squares Reported as Entries in this Table	Total Sums of Squares of Elements in Residual Matrices						
		ORTHOGONAL Transformations				ORTHONORMAL Transformations		
		Communalities in Diagonals of R_y		Ones in Diagonals of R_y		Ones in Diagonals of R_y		Communalities in Diagonals of R_y
		T_o	T_p	T_o	T_p	T_n	T_m	T_m
$(F_{yr} - F_x)$	32 x 8 = 256	1.3258	1.3667	1.8747	1.9054	.2133	.1790	1.4270
$(F_{yr}F'_{yr} - R_y)$, other than diagonal elements	32 x 31 = 992	.6612	.6612	2.0005	2.0005	3.5448	3.5448	4.4173
$(F_{zr}F'_{yr} - R_{zy})$	32 x 5 = 160	.2162	.2162	.1965	.1965	.3405	.3405	.4163
$(F_{zr}F'_{zr} - R_z)$	$(5)^2 = 25$	8.898	8.898	9.1654	9.1654	9.8459	9.8459	8.0030

SOME LEAST-SQUARES TRANSFORMATIONS OF REGRESSION ESTIMATORS OF ORTHOGONAL FACTORS

Emil F. Heermann
U. S. Army Personnel Research Office

This paper provides a brief introduction to the factor analysis model used by psychologists and presents some methods for transforming regression estimators of orthogonal factors to a more useful form.

I. THE FACTOR ANALYSIS MODEL. Factor analysis is a mathematical method for representing n correlated measures obtained on N individuals by means of a linear combination of hypothetical measures, called factors. Psychologists draw a distinction between common factors, which are related to two or more measures, and specific factors which are related to only one of the original measures. In general, then, the linear model postulates n specific factors and r common factors where r is less than the number of measures n. The sampling theory for such a model is not yet thoroughly understood and this paper will confine itself to algebraic rather than statistical considerations.

The linear model of factor analysis can be most conveniently presented in the form of matrix equations:

(1) $$Z = Z_c F' + Z_s S' \text{ , where}$$

$Z = N \times n$ matrix of observed measures (zero mean, unit variances)

$Z_c = N \times r$ matrix of common factor scores

$F' = r \times N$ matrix of common factor weights

$Z_s = N \times n$ matrix of specific factor scores

$S' = n \times n$ diagonal matrix of specific factor weights

Terms on the right side of the equation are all unknown. In (1) it is assumed that:

(2) $$\frac{1}{N} Z'_c Z_c = I_r$$

(3) $$\frac{1}{N} Z'_s Z_s = I_n$$

(4) $$\frac{1}{N} Z'_c Z_s = 0$$

These assumptions allow us to find F' and S' from the observed measures, but before we examine this procedure it is important to indicate a rather serious difficulty with our linear model. When we postulate $n + r$ factors to explain n measures this invariably means that there exist multiple solutions for an individual's factor scores. Solutions exist because the linear equations are consistent but the rank of the coefficient matrix does not allow unique solutions. To get around this difficulty psychologists have utilized least-squares procedures for estimating the factor scores from the observed measures, and we shall examine the defects of these estimators at a later point.

We now return to the problem of finding F' and S' from the observed data. Because of (2), (3) and (4) we may write:

$$R = \frac{1}{N} Z'Z = FF' + SS', \text{ or}$$

$$R - SS' = FF'$$

Since SS' is a diagonal matrix, we begin by estimating values for SS' and then obtain the eigenvectors and eigenvalues of

$R - SS' = A\,K\,A'$, where

A = $n \times r$ matrix of eigenvectors

K = $r \times r$ matrix of eigenvalues

Now, $F = AK^{1/2}$, and the sums of squares of the rows of F are used to obtain new values of SS', and the process is continued until $R - SS'$ is fit with the minimum rank F. This whole process may be thought of as finding a set of values for SS', such that $R - SS'$ is of minimum rank.

It will be noted that F is arbitrary up to an orthogonal transformation and it is necessary to postulate some method for finding a psychologically meaningful F. Psychologists follow L. L. Thurstone here and attempt to transform the arbitrary F, so that it approximates "simple structure", i.e., has a maximal number of zero or near-zero entries. To accomplish this we find some orthogonal transformation, λ, such that

$$F\lambda = F_R$$

where $\lambda'\lambda = I$, and

F_r approximates "simple structure". A considerable number of computer programs exist which determine by analytical means the best transformation λ.

II. ESTIMATION OF FACTOR SCORES. Assuming now that F_r is adequately fixed, our problem is then to find the values of Z_c and Z_s that will satisfy the linear model specified in (1). Since an infinite set of such factor scores exist, psychologists commonly turn to a regression method for estimating Z_c for a fixed Z_s. We are primarily interested here in finding values for Z_c, and the least-squares solution for Z_c is easily found to be

$$ZR^{-1}F_R = Z_c .$$

But unlike the factor scores Z_c, these estimators are intercorrelated, because

$$(5) \qquad \frac{1}{N}\hat{Z}'_c\hat{Z}_c = F'_R R^{-1} F_R, \quad \text{the}$$

covariance matrix of the least-squares estimator is not diagonal. Adjusting the covariance matrix so that we have an intercorrelation matrix gives:

$$R_L = D_e^{-1} F'_R R^{-1} F_R D_e^{-1},$$

where

$D_e = r \times r$ diagonal matrix, formed from the square roots of diagonal entries of (5), the covariance matrix.

A further practical difficulty is that regression estimators are not univocal, i.e., each estimator is correlated with more than one common factor. This can be seen by examining the matrix of correlations between the factor scores and the least-squares estimators, which is

$$R_{fL} = D_e R_L.$$

It is not too difficult to derive transformations of the beta weights used in the regression estimators that remove these defects, but it is not possible by means of a single transformation to simultaneously remove both defects.

To adjust for non-orthogonality, we first find an arbitrary set of orthogonal vectors that serves as a vector basis for the regression estimators. To do this we find an $r \times r$ T, such that

$$R_L = TT'$$

The elements of T represent the correlations of the least-squares estimators with the r orthogonal axes. The correlations of the factors with the orthogonal axes are given by

$$R_{fT} = D_e T,$$

and to find the best set of orthogonal axes we must transform T so that $D_e T$ most closely approximates the diagonal matrix D_e. In matrix algebra, we wish to find a λ, such that if

$$R_{fT}\lambda - D_e = E, \text{ then}$$

$$\text{trace } E'E = \text{minimum, and}$$

$$\lambda' \lambda = I.$$

From a theorem proved by B. F. Green (1) we find

$$\lambda = (T'D_e^4T)^{-1/2}\, T'\, D_e^2 \ .$$

Now if the requirements of the psychologist are such that it is more important to have univocal rather than orthogonal estimators than a somewhat different procedure is required. In this case we determine so that if

$$R_{fT}\lambda - D_e = E \ ,$$

$$\text{trace } E'E = \text{minimum, and}$$

$$\lambda'_{.i}\lambda_{.i} = 1, \ (i = 1, 2, \ldots r), \text{ where}$$

$$\lambda_{.i} = i^{th} \text{ column of } \lambda.$$

As can be seen the orthogonality restraint on λ has been relaxed. The normal equations for these conditions are

$$(T'D_e^2T - \gamma)\lambda = T'D_e^2 \ , \tag{8}$$

where γ = r x r matrix of Lagrangian Multipliers. The above equation does not allow us to find a matrix expression for λ which does not also involve γ. Hence we pursue a simpler approach which involves a least-squares solution for λ followed by imposing the restraints that each column of λ have sums of squares equal to unity. With this approach we find that

$$\lambda_n = T^{-1}D, \text{ where}$$

D is the diagonal matrix of constants needed to adjust T^{-1} so that the column sums of squares equal unity. When the least-squares estimators are transformed by λ_N we find that the matrix of correlations between Z_c and our transformed estimators is $R_{fN} = D_eD$, which is quite obviously a diagonal matrix.

III. AN APPLICATION OF THE TRANSFORMATIONS. In 1954 APRO carried out a factor analysis of visual acuity tests administered during dark adaption. The examinees in this experiment were 100 soldiers from Fort Myer, Virginia. The design of the experiment called for pre-adapting subjects to a high brightness level and then during the period of dark adaption the examinees were tested on visual acuity targets presented at scotopic, mesopic, and low photopic brightness. The various acuity tests (Modified Landolt Ring and Chevron Contrast Adoption tests) were presented in a modified Armed Forces Vision Test.

The 35 variable intercorrelation matrix obtained in this experiment yielded 8 orthogonal factors which were rotated to orthogonal simple structure. To illustrate the derivations presented in this paper, a submatrix was selected from the 35 x 8 complete factor matrix. Table 1 gives this submatrix.

Table 1

Illustrative Factor Matrix

Variable Name		I	II	III	IV
Landolt Scotopic	1	.84	.07	.13	.34
Landolt Low Photopic	2	.16	.85	.03	.13
Chevron Scotopic	3	.28	.32	.12	.75
Chevron Mesopic	4	-.01	.03	.79	.29

Factor I was interpreted to represent "Rod-Adapted Resolution"; Factor II was interpreted to represent "Cone Adapted Resolution"; Factor III was interpreted to represent "Cone Adapted Brightness Discrimination", and Factor IV was interpreted as "Rod-Adapted Brightness Discrimination".

Table 2 compares the classic regression equations approach to estimating these four factors with the methods derived in this paper. As noted previously the least-squares estimators are intercorrelated and fail to be univocal. The univocal estimators achieve an ideal pattern of correlations with the factors but they are intercorrelated somewhat more than the least-squares estimators. The orthogonal estimators, while uncorrelated, are not univocal although they are closer to being univocal than the least-squares estimators. Inevitably then, choice of any one solution means that certain defects must be tolerated. From the results

presented in Table 2 it would appear that the orthogonal estimators represent something of a compromise between the maximum validity of the least-squares estimators and purity of the univocal estimators.

---

(1) Green, B. F. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, 1952, 17 429-440.

A RELIABILITY TEST METHOD FOR "ONE-SHOT" ITEMS

H. J. Langlie
Aeronutronic Division, Ford Motor Company

1. INTRODUCTION. As a result of many reliability problems which have plagued procuring agencies in the missile and space programs, increased emphasis has been placed on the development of improved reliability demonstration test methods. In this connection, the Army has advocated increased use of the "test-to-failure" concept to establish the existence of satisfactory margins of reliability with respect to critical factors. This philosophy has the advantage in that statistical statements can be made regarding reliability on the basis of relatively small samples. This paper will discuss the application of this general concept to a particular class of hardware.

In essence, the "test-to-failure" concept involves submitting a test specimen to an increasing environment or load stress until failure is detected. By observing the statistical behavior of the stresses at which failure occurs, the lower limit of stress below which the probability of failure is very small can be selected by using the mean and standard deviation of these data. Robert Lusser (reference 1) advocated the safety margin concept in interpreting these data. As is shown in Figure 1, the larger the value of k (which is the distance from the mean strength to the upper limit of the operating applied stress divided by the standard deviation of the data), the greater the reliability of the specimen with respect to the stress involved.

In establishing the reliability objectives for the Shillelagh Program, the Army Missile Command required that safety margins be demonstrated in the test laboratory for Shillelagh components with respect to critical environmental stress factors. Many of these components are of a "go-no-go" type such as thermal batteries, electrical relays, and other short-lived equipment items. In most cases, little information was available prior to test regarding the nature of the standard deviation of the distribution of strenghts for these parts. Furthermore, it was desired to perform a laboratory test involving a minimum number of samples. A review of attribute sensitivity testing techniques such as the Up and Down method (reference 2) and the Probit method (reference 3) indicated that these methods cannot be applied satisfactorily under the sample size and technical limitations imposed. As a result, a study was made to develop a method for selecting stress levels for testing which required no a'priori assumption regarding the standard deviation of the unknown strength distribution and could be performed satisfactorily with sample sizes of the order of fifteen or twenty.

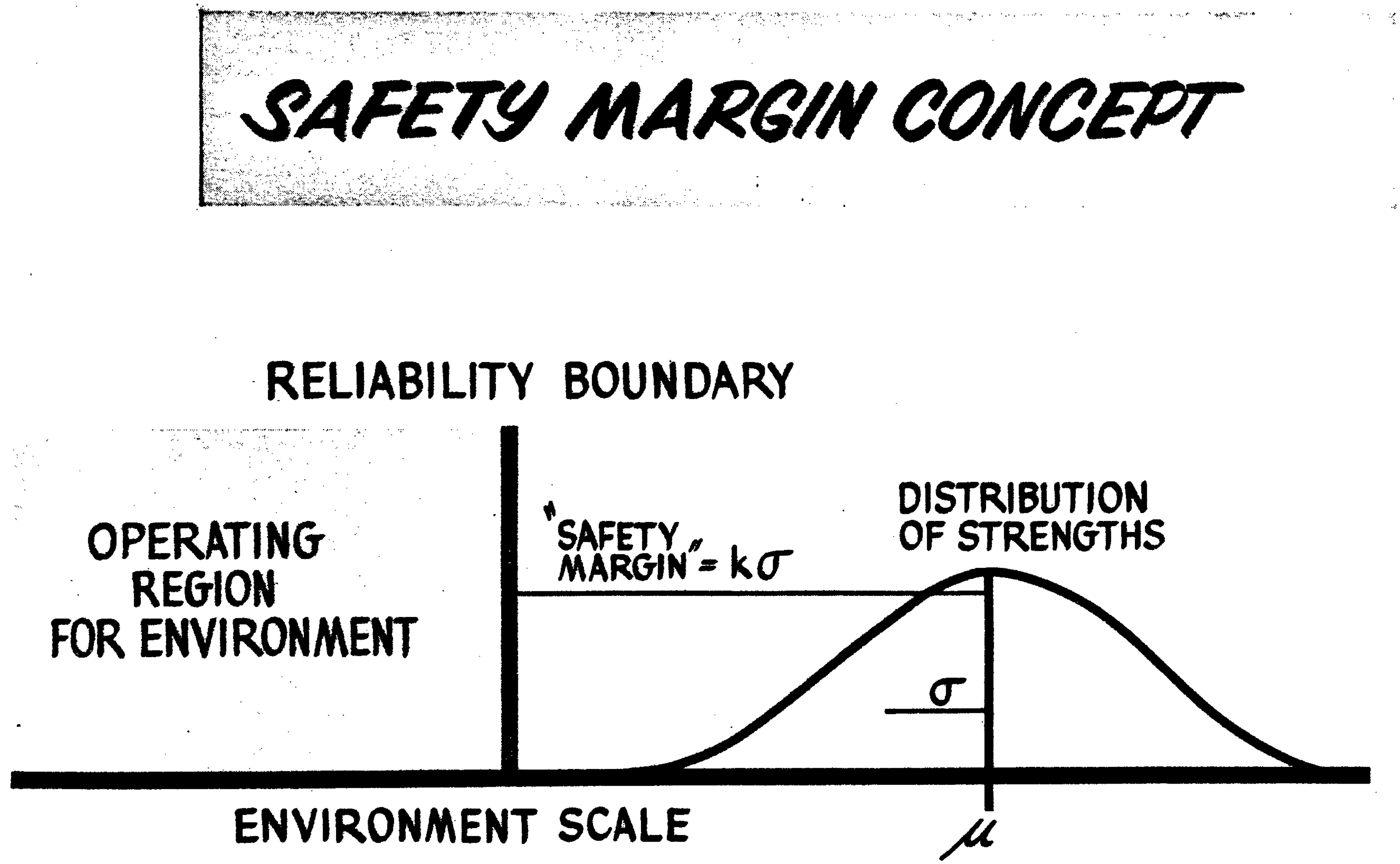

SAFETY MARGIN CONCEPT
RELIABILITY BOUNDARY
OPERATING REGION FOR ENVIRONMENT
"SAFETY MARGIN" = kσ
DISTRIBUTION OF STRENGTHS
σ
ENVIRONMENT SCALE
μ

FIGURE 1.

Detailed empirical investigations were performed, using Monte Carlo methods with a high speed digital computer, to develop a satisfactory algorithm for determining successive stress test levels as the experiment proceeds. Exact maximum likelihood equations were used to calculate the statistical estimates, μ_e, and σ_e, of the population parameters, μ and σ, of the distribution of part strengths. By repeated simulations of experiments for sample sizes ranging from ten to 900, empirical curves were obtained showing the variance in the calculated estimates vs sample size. After this study was complete, a technical report was prepared (reference 4). The paper today will discuss some significant results contained in this report.

Since the completion of the study, several applications have been made in reliability testing electrical and mechanical components. One such application is presented in this paper by way of illustration.

2. DISCUSSION. In order to proceed with the discussion of the test method, a specific definition is given regarding the terms "stress" and "strength" as follows:

> Stress is a test factor, such as environmental level or force level which is applied to the test specimen. Operational stresses represent the mix of environmental or load conditions that can be expected to be imposed on a typical specimen during its life. During the conduct of a "one-shot" test, the stress represents the applied test factor which is varied in magnitude from specimen to specimen in a systematic manner.
>
> Strength is a property ascribed to a specimen such that if the stress imposed on the part is greater than the strength of the part, the part will fail. Conversely, if the stress is less than the strength of the part, the part will not fail.

Failure in the sense used above is a general term referring to unsatisfactory completion of function, out of tolerance performance, breakage, or other evidence of malfunction. For each test attempt wherein a stress is applied to a specimen, there is associated an outcome which is a binary variable: success or failure.

It is assumed that, given a homogenous sample of replicate specimens, the part strengths are distributed normally with an unknown mean and

standard deviation. The purpose of the test method is to select stress levels in such a way as to generate outcomes which can then be used to calculate statistical estimates of the parameters of the distribution of part strengths.

3. PERFORMING THE TEST. In undertaking to perform the "one-shot" test, there are three steps to be taken which should be followed regardless of the nature of the application. They are:

1. Establish the criteria of failure, or acceptance.
2. Determine the test interval.
3. Select the stress levels.

The last of these three steps, selecting the stress levels, proceeds concurrent with the actual performance of the test. These three steps are discussed below.

3.1 Establishing Failure Criteria.

It is very important that painstaking care be given to the set of ground rules which will be followed for differentiating between a success or a failure. For purposes of reliability testing, this means that a careful enumeration should be made of all undesirable responses of a test specimen, such as particular modes of failure, out of tolerance performance, and any other mode of unacceptable product performance. These criteria are the very basis for establishing product assurance in the laboratory and should be reviewed and approved by all qualified parties having a technical interest in the product.

3.2 Determining the Test Interval.

In order to proceed with a generation of stress levels for testing purposes, it is necessary to choose a test interval which is used as a basis for the stress sequencing method. This interval should be selected large enough to include all possible ranges of strengths of the parts to be tested. This interval can be made conservatively large, since the "one-shot" method has been designed to cause the stress levels to be generated in the vicinity of interest (i.e., in the vicinity of the distribution of strengths) as the test proceeds. As a sample illustration, the range for a drop height test for glass containers designed to withstand say, a six inch drop, could be chosen to have a lower limit of zero and an upper limit of three feet. The method of analysis of the data is such that the particular choice of the

endpoints of the test interval do not have an appreciable effect on the results for sample sizes of fifteen or more. In the event that the test interval turns out, as the test proceeds, to be inappropriately chosen, then the stress levels will tend to converge towards one limit or the other. In such an event, particularly in the case of reliability testing, convergence towards the lower level is usually indicative of a totally unsatisfactory product, whereas convergence towards the upper limit can be shown to be statistically acceptable by use of the likelihood ratio test.

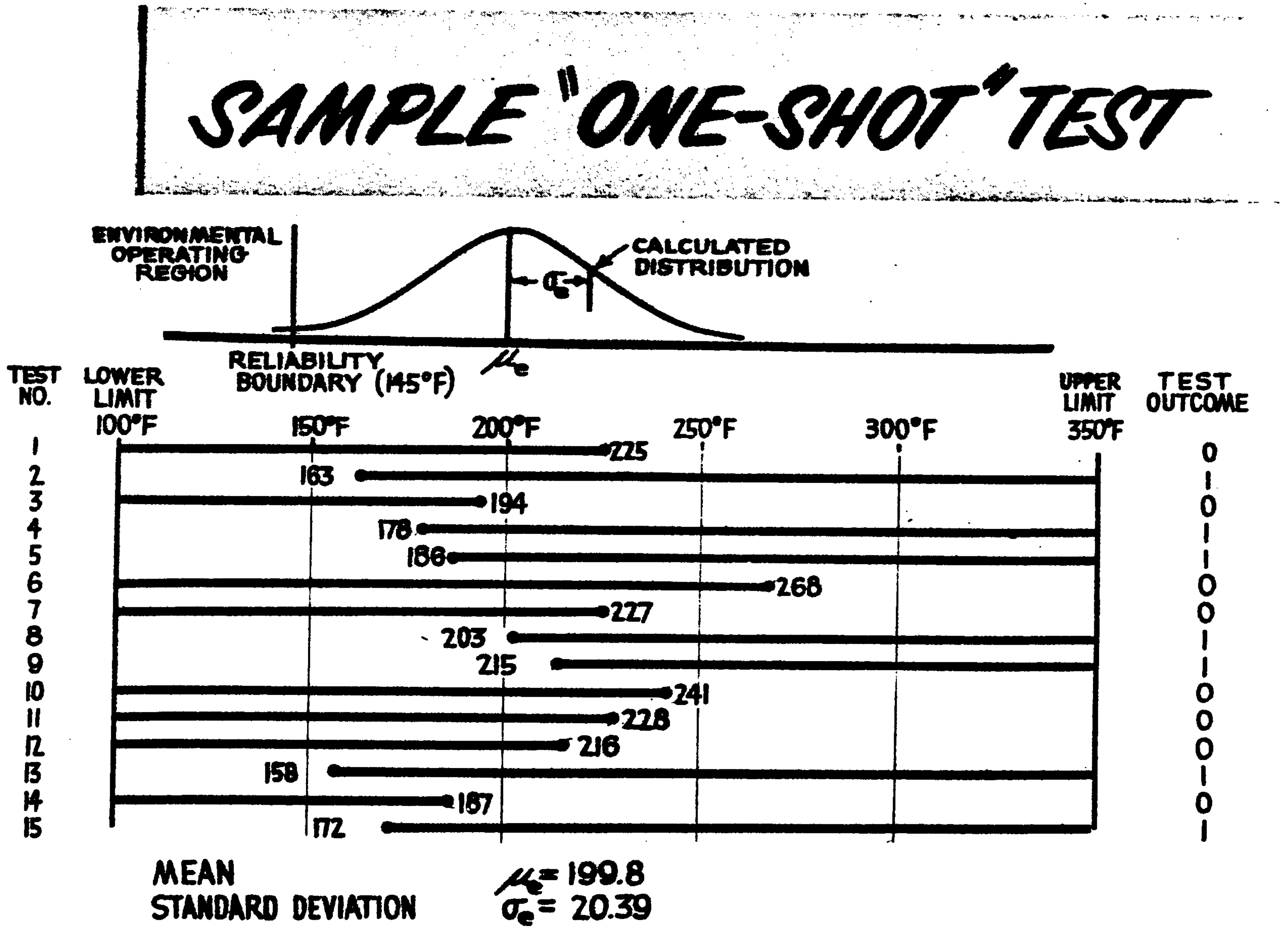
In Figure 2 there is represented the results of an actual "one-shot" test on thermal batteries to determine the reliability with regard to high temperature. In this instance, the batteries were designed to perform reliably at 145°F. On the basis of conservative engineering judgement and some limited development test data, the lower limit was selected to be 100°F (the level at which all thermal batteries would be expected to perform satisfactorily) and the higher limit was selected to be 350°F (the level at which all thermal batteries would be expected to fail).

3.3 Selecting the Stress Levels.

Once the test interval and failure criteria have been established, the test commences by selecting the first stress level at the midpoint of the interval. After exposing the first specimen to this environmental level and activating it, a one or zero is recorded to indicate the outcome as a success or failure respectively (see Figure 2).

The general rule for obtaining the $(n + 1)^{st}$ stress level, having completed n trials, is to work backward in the test sequence, starting at the n^{th} trial, until a previous trial (call is the p^{th} trial) is found such that there are as many successes as failures in the p^{th} through the n^{th} trials. The $(n + 1)^{st}$ stress level is then obtained by averaging the n^{th} stress level with the p^{th} stress level. If there exists no previous stress level satisfying the requirement stated above, then the $(n + 1)^{st}$ stress level is obtained by averaging the n^{th} stress level with the lower or upper stress limits of the test interval according to whether the n^{th} result was a failure or a success.

To illustrate, suppose it is desired to find the second stress level in Figure 2. Since there was only one previous observation (i.e., first unit failed) it is not possible to find a stress level where all intervening results even out. That is, the second stress level is obtained by averaging the first with the lower limit. To find the eighth stress level, it is observed that results from test 4 through 7 (i.e., the last four results) cancel each other out. Thus, the eighth stress level is obtained by averaging the fourth.



TEST-TO-FAILURE OF THERMAL BATTERIES IN TEMPERATURE

FIGURE 2.

stress level with the seventh.

As a final example, it is observed that after the twelfth test has been completed, there again exists no previous stress level for which the number of failures equals the number of successes. Since the twelfth test was a failure, the thirteenth stress level is obtained by averaging the twelfth stress level with the lower limit.

As an aid in identifying the important parameters of the test, the stress level is designated by the letter s and the outcome is designated by the letter u. The lower limit of the test interval is designated A and the upper limit is designated B. Upon the conclusion of the test, the stress values, $(s_1, s_2, \ldots, s_N)$, and the corresponding outcomes, $(u_1, u_2, \ldots, u_N)$, where N equals the test sample size, are used to perform a complete analysis for hardware reliability.

4. DERIVATION OF MAXIMUM LIKELIHOOD EQUATIONS. Consider the random sample $X = (x_1, x_2, \ldots, x_N)$ of N observations where the x_i are independent random variable from a Gaussian distribution, $g(x;\mu, \sigma)$, with mean μ and standard deviation σ. Consider also an N-dimensional vector $S = (s_1, s_2, \ldots s_N)$ where $A \leq s_i \leq B$. From this, construct a third vector $U = (u_1, u_2, \ldots, u_N)$ where

$$u_i = 1 \text{ if } s_i < x_i$$
$$u_i = 0 \text{ if } s_i \geq x_i$$

The variable x_i is called the strength of the i^{th} part; s_i is called the applied stress level for the i^{th} part, and u_i is called the outcome of the "test" on the i^{th} part. The outcome $u = 1$ is called a success (i. e., the applied stress was less than the part strength) and, conversely, $u = 0$ is called a failure.

The object of this development is to obtain formulas for calculating the estimates μ_e and σ_e of μ and σ, given only S and U. This will be accomplished by obtaining values of μ_e and σ_e which maximize the likelihood (i. e., probability) of obtaining the outcome U given S.

The probability of outcome u_i given s_i can be written

(1) $$p_i = \text{Prob}\left[u_i \middle| s_i\right] = u_i \int_{s_i}^{\infty} g(v;\mu,\sigma)dv + (1-u_i)\int_{-\infty}^{s_i} g(v;\mu,\sigma)dv$$

The probability of outcome U can be written as the product of the probabilities of the individual u_i since the x_i are independent.

(2) $$P\left[U\right] = \prod_{i=1}^{N} p_i = L(\mu, \sigma)$$

The expression $P\left[U\right]$, when regarded as a function of the population parameters μ and σ, becomes the likelihood function for outcome U (ref. 5).

To find the values of μ and σ (now regarded as variables) which maximize L, we differentiate (2) with respect to μ and σ and solve the system

(3) $$\partial \ln L/\partial \mu = 0$$
$$\partial \ln L/\partial \mu = 0$$

where the logarithm of L is used to simplify the algebra.

Letting $t_i = (s_i - \mu)/\sigma$, $g_o(v) = (2\pi)^{-1/2} \exp(-v^2/2)$, the normalized Gaussian, and remembering that u_i can take on only values of 0 or 1, equation (1) can be re-written

(4) $$\ln p_i = u_i \ln\left[\int_{t_i}^{\infty} g_o(v)dv\right] + (1-u_i)\ln\left[\int_{-\infty}^{t_i} g_o(v)dv\right]$$

The following definitions ($\stackrel{d}{=}$) and derivations will be helpful:

(5) $$G(t) \stackrel{d}{=} \int_{-\infty}^{t} g_o(v)\, dv$$

(6) $$dg_o(t)/dt = -t g_o(t)$$

(7) $$\partial t/\partial \mu = -1/\sigma \qquad (\sigma > 0)$$

(8) $$\partial t/\partial \sigma = -t/\sigma \qquad (\sigma > 0)$$

Since $\ln L = \Sigma \ln p_i$, equations (3) become

(9) $$\partial \ln L/\partial \mu = \Sigma\, \partial \ln p_i/\partial \mu = 0$$
$$\partial \ln L/\partial \sigma = \Sigma\, \partial \ln p_i/\partial \sigma = 0$$

From (4)

$$\text{(10)} \qquad \partial \ln p_i/\partial\mu = \frac{-u_i g_o(t_i)}{1 - G(t_i)} \frac{\partial t_i}{\partial\mu} + \frac{(1-u_i)g_o(t_i)}{G(t_i)} \frac{\partial t_i}{\partial\mu}$$

$$= \frac{g_i}{\sigma}\left[\frac{u_i}{1-G_i} - \frac{1-u_i}{G_i}\right]$$

where the arguments (and subscript "o") have been omitted to simplify writing. Similarly,

$$\text{(11)} \qquad \partial \ln p_i/\partial\sigma = \frac{t_i g_i}{\sigma}\left[\frac{u_i}{1-G_i} - \frac{1-u_i}{G_i}\right]$$

Denoting by h_i the expression in brackets and eliminating the constant σ, equations (9) become

$$\left.\begin{aligned} \partial \ln L/\partial\mu &\overset{d}{=} p(\mu, \sigma) = \sum_{i=1}^{N} g_i h_i = 0 \\ \partial \ln L/\partial\sigma &\overset{d}{=} q(\mu, \sigma) = \sum_{i=1}^{N} t_i g_i h_i = 0 \end{aligned}\right\}$$

Equations (12) are valid only if the value of σ which satisfies the maximum likelihood equations is non-zero. A quick examination of the data can be made to determine if a non-zero σ is a maximum likelihood solution. If the maximum stress level at which a success occurred is greater than the minimum level at which a failure occurred, than a non-zero σ satisfies equations (12). If this statement is not true, then $\sigma = 0$ represents a maximum likelihood estimate for the standard deviation and the maximum likelihood estimate of the mean is a connected interval contained between the maximum failure stress level and the minimum success stress level which represent the lower and upper bounds of the interval respectively. The latter situation illustrates an outcome which in fact must be achieved if all of the part strengths were concentrated at a mass point within the above mentioned interval. The maximum likelihood corresponding to this outcome is one. Although unique estimates cannot be obtained for μ_e and σ_e in such an instance, it is possible to provide a basis for decision-making, using the likelihood ratio statistic along with a suitably

constructed hypothesis (reference 4). A result of this type is referred to as degenerate. It should be mentioned that, for a fixed population standard deviation, $\sigma \neq 0$, the probability of obtaining a degenerate result approaches 0 as the sample size becomes large. The proof of these statements is contained in reference 4 and is beyond the scope of this paper.

5. STATISTICAL PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATES. The preceding sections describe the method for selecting stress levels and the likelihood equations for calculating μ_e and σ_e. In order to make practical applications of the method, however, information is required regarding the statistical properties of the maximum likelihood estimates corresponding to various sample sizes. To facilitate obtaining this information, an extensive computer simulation was carried out using Monte Carlo techniques. In this way, hundreds of values of μ_e and σ_e could be obtained corresponding to hundreds of simulated experiments using random numbers for part strengths. Statistical summaries were then obtained and the variance of the estimates of the parameters empirically derived.

To perform the simulation, a standard interval, A = 1, B = 1, was chosen. The sampling of strengths was simulated by converting the sum of twelve two-digit numbers, constructed from a file of one millions random digits, to a random deviate with population mean μ and standard deviation, σ. Two populations were employed: $\mu = 0$ and $\sigma = 0..25$, and $\mu = 0.2$ and $\sigma = 0.1$. For each population, one hundred runs, each consisting of N samples, were made for N = 4 through 15, 20, 25, 30 and 35, with an additional four hundred runs for N = 15 and N = 30 for additional information on the distributions of the estimates. Finally, four runs of N = 900 were made for each population to empirically investigate the asymptotic convergence of (μ_e, σ_e) to (μ, σ).

For each set of 100 runs at a fixed sample size, the mean and variance were averaged separately and plotted as shown in Figures 3 and 4. Straight lines were then fitted to the data, recognizing that some spurious effects are introduced for small sample sizes (i. e., 15 or less) due to the discreteness of the admissible outcomes which are possible. (For sample size of N, only 2^N outcomes are possible corresponding to the 2^N possible configurations of 0 and 1.)

Based on the results of Figures 3 and 4, the variance of the estimates are given by:

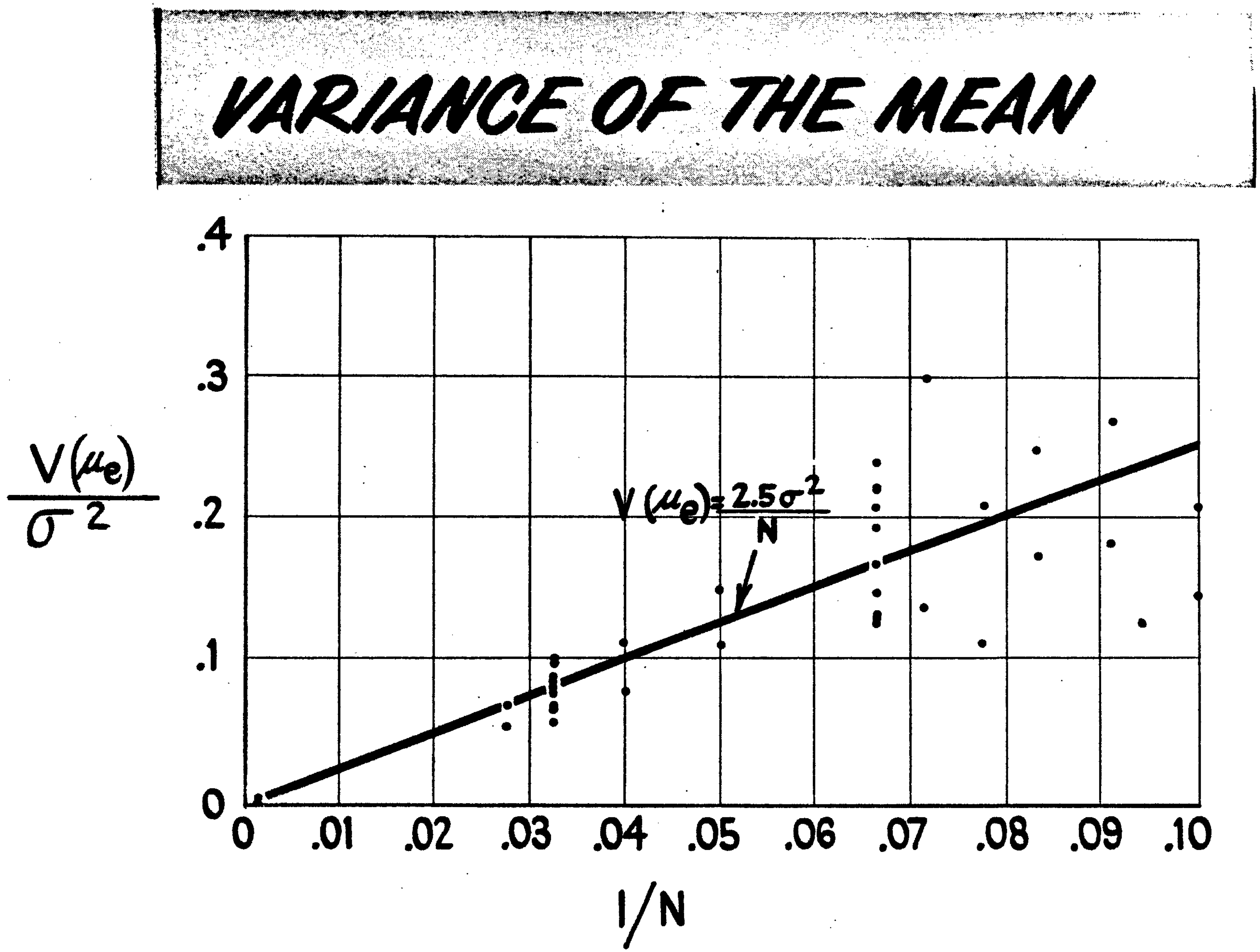

VARIANCE OF THE MEAN
$\frac{V(\mu_e)}{\sigma^2}$
$V(\mu_e) = \frac{2.5\sigma^2}{N}$
.4
.3
.2
.1
0
0
.01
.02
.03
.04
.05
.06
.07
.08
.09
.10
1/N

FIGURE 3.

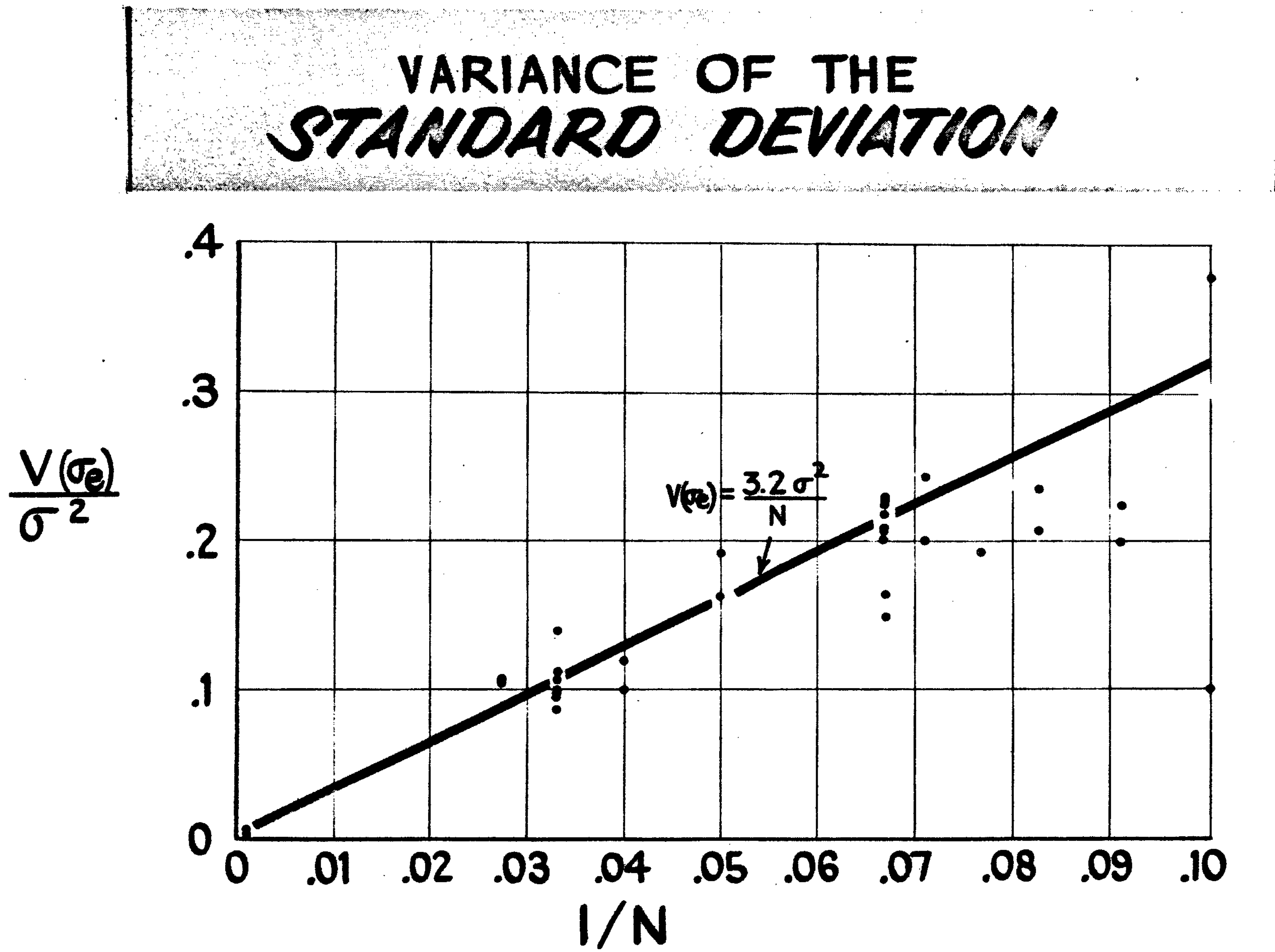

VARIANCE OF THE
STANDARD DEVIATION
$\frac{V(\sigma_e)}{\sigma^2}$
.4
.3
.2
.1
0
$V(\sigma_e) = \frac{3.2\sigma^2}{N}$
0
.01
.02
.03
.04
.05
.06
.07
.08
.09
.10
1/N

FIGURE 4.

$$V(\mu_e) = 2.5\sigma^2/N$$

$$V(\sigma_e) = 3.2\sigma^2/N$$

In order to use these formulas in practical applications, the unbiased estimate of σ, denoted $\hat{\sigma}$, is substituted for σ in the above formulas. The relationship between the unbiased estimate and the maximum likelihood estimate is given empirically in Figure 5. The variance of the unbiased estimate can be calculated using the relation

$$V(\hat{\sigma}) = \frac{V(\sigma_e)}{\beta^2} = \frac{3.2\hat{\sigma}^2}{N\beta^2} = \frac{3.2\sigma_e^2}{N\beta^4}$$

The above formula is sufficiently accurate for sample sizes on the order of 50 or greater, wherein the distribution of the estimate of the standard deviation approaches the normal distribution. For smaller size samples it was observed from the empirical study that $n\hat{\sigma}/\sigma$ approximately follows the chi-square distribution with n degress of freedom where n is given by

$$n = 0.625\beta^2 N \text{ (reference 4)}$$

where N is the sample size and β is given by Figure 5.

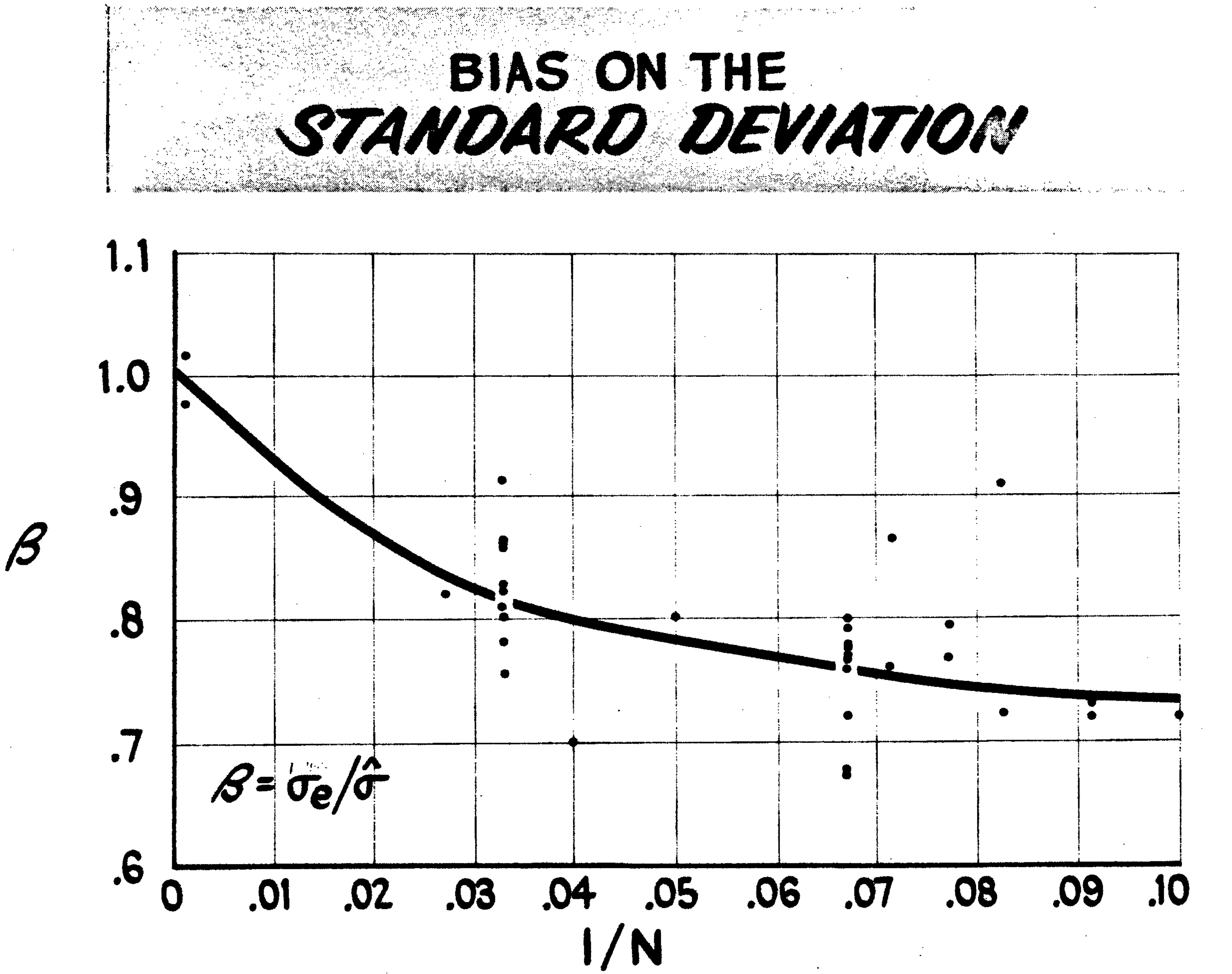

BIAS ON THE
STANDARD DEVIATION
β
1.1
1.0
.9
.8
.7
.6
$\beta = \sigma_e/\hat{\sigma}$
0
.01
.02
.03
.04
.05
.06
.07
.08
.09
.10
1/N

FIGURE 5.

REFERENCES

1. Lusser, R., Reliability Through Safety Margin, Research and Development Division, Army Rocket and Guided Missile Agency, Redstone Arsenal, Alabama, October 1958.

2. Dixon and Massey, Introduction to Statistical Analysis, McGraw-Hill 2nd Edition, 1957, Chapter 6.

3. Finney, D. J., Probit Analysis, Cambridge University Press

4. Langlie, H. J., A Reliability Test Method for "One-Shot" Items, Aeronutronic Division of Ford Motor Company Publication No. U-1792, August 10, 1962.

5. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, 6th printing, 1954, p. 478.

INVESTIGATION IN TEMPERATURE CONTROL OF HYDRAULIC SYSTEMS IN ROUGH TERRAIN FORK TRUCKS

Irving Tarlow
Quartermaster Research and Engineering Command
Natick, Massachusetts

The problem of overheating in hydraulic systems of mobile equipment has not been given full consideration nor have the ramifications been completely understood by designers of such equipment. The effect of excessive heat can be injurious to the operator, the equipment and to persons in the proximity of the equipment.

In industrial applications of hydraulic equipment, space has not been one of the problems. Hydraulic systems designed on this premise provided for a reservoir tank of sufficient fluid capacity to dissipate the heat or a system in which a heat exchanger has been incorporated.

This approach cannot be applied to mobile hydraulic systems. Space is not available for large fluid reservoirs or heat exchangers. The situation is usually more aggravated by the radiation of heat from the engine as well as crowding components of the system into a small enclosed area in which the heat from the system itself has to be considered.

The designer, evaluating the heat effect, takes into consideration the type of operation for which the equipment was intended. For commercial materials handling equipment this approach has been for the most part valid. The hydraulic system is not taxed to the degree that heat has been a factor. Idle time between cycles has been sufficient to allow cooling of the fluid to a safe temperature level. However, in Military applications, where unloading operations are on a 24-hour per day basis and long road marches are a required operational feature, temperature control must be built in.

The condition of high fluid temperature was detected in an All Purpose Fork Lift Truck having a 6,000-lb capacity at 24-inch load center. This vehicle is a Military item presently in the supply system in large quantities. The purpose of this vehicle is to handle supplies, whether palletized, unitized or containerized, over rough terrain, in open storage areas and in all types of climatic environments as well as all types of terrain. This vehicle must be capable of operating under conditions of limited visibility, inclement weather and blackout and be capable of

negotiating open rolling and hilly terrain consisting of mud, snow and sand. In addition, this vehicle must also be capable of highway operations.

Prior to actual investigation and subsequent remedial action, the possible effects of the problem were analyzed. The analysis included but was not limited to:

(1) The effect on humans.

(2) The effect on the system.

The location of the operator's compartment is directly above the reservoir tank. Therefore, operator discomfort is a factor. Since the tank is mounted outboard of the chassis, personnel coming in contact with the tank surface could suffer burns. While either and both of the above are serious, far more important were the injuries and possible fatalities which could result from hydraulic line ruptures and component failures. The unleashing of hydraulic fluid which was under pressure of up to 1400 psi and whose temperature was in excess of 250 F is considered to be the most serious of those previously mentioned.

The second area to be considered in evaluating the effect of high temperature rise is the effect upon the system itself. Excessive temperatures of petroleum fluids will cause lubricating qualities to be reduced. For gear pumps whose shafts run in journal bearings, the design is such that the shaft "floats" on a film of oil so that there is no contact between the shaft and the bearings. Extreme heat will cause the film to break down. When this occurs, the shaft comes in contact with the bearing and causes galling and seizing. Galling and seizing will also occur when unequal expansion occurs between two parts of dissimilar material. Directional control valves are particularly susceptible to the problem of unequal expansion which causes spools to stick resulting in an inoperative system.

Petroleum fluid deterioration can have additional harmful effects upon the system. When a petroleum fluid deteriorates sludge forms, this sludge will be deposited on surfaces which are subjected to localized heating. Filters and lines become clogged and valves become stuck. If a sump filter is used in the reservoir tank and it should become clogged due to the formation of sludge, cavitation of the pump will in all likelihood result.

The particular hydraulic system on which the investigation was conducted is of the open center type and consists of a fixed displacement pump coupled to the engine input shaft. Fluid passes from the pump to a directional control valve and is then directed either to reservoir tank when no work is required or to a working cylinder.

It was necessary to first isolate the cause of excessive heating. A work cycle was devised based on a simulated field operation. This consisted of lifting a 6,000-lb rated load, transporting it to a new site, stacking the load and then returning to the starting position. Temperature was recorded in the reservoir tank approximately 2 inches from the bottom and 18 inches away from suction outlet.

Through observation, it was detected that the rise in temperature occurred during the transporting phase of the cycle. Temperature would drop during the work cycle and then rise again during transport. By analyzing the circuitry, it was determined that the heat rise was in all probability due to a restriction and further that this restriction was located in the directional control valve.

In order to substantiate the analysis, the directional control valve was by-passed and the fluid path was controlled so that it passed from the pump directly to the reservoir. This resulted in a substantial reduction in temperature and stabilization point in the system thereby validating the analysis.

The problem of controlling the fluid temperature became twofold:

(1) A method had to be devised by which vehicles already issued to the field could be modified quickly and economically and still permit satisfactory operation during the service life of the vehicle. (Service life is considered to last from 8 to 11 years for this type of equipment.)

(2) Redesign of the hydraulic system in order to produce specification changes which would permit future procurement of a vehicle capable of satisfying all of the operational requirements.

Prior to development of a solution, it was first necessary to establish acceptable operating parameters. To do this, the vehicle was instrumented to record reservoir tank temperature, pressure between pump and directional control valve and pressure on the return line between the control valve and reservoir tank.

Since the pump was of a fixed-displacement type directly coupled to the engine, the controlled variable became engine speed. The pump output in gallons per minute was determined at various engine speeds. After this, the back pressure in pounds per square inch at various flow rates (gpm) was determined for the directional control valve. The next step was to determine the acceptable back pressure and relative temperature rise limits. This was necessary since it already had been determined that 100 deg. temperature rise above ambient could be tolerated for the particular oil being used in the system and for areas of the world where the vehicle would be used. In order to ascertain, therefore, what back-pressure could be tolerated which would result in no more than 100 deg. temperature rise, the following tests were conducted:

The vehicle was operated at maximum engine speed until the temperature in the tank stabilized for a minimum period of 20 minutes. Engine speed was varied and successive runs at different speeds were made. Through observation and analysis of the stabilized temperature and back pressure, it was determined that the naximum open center pressure drops across the directional control valve should not exceed 100 psig.

In explanation of the above let me, at this time, cite a few of the specifics:

a. The majority of test runs were made in low gear and at full engine rpm. This will reproduce all conditions anticipated during convoy with the exception of the cooling effect from a high velocity wind passing by the hydraulic tank. When tested in high gear at 25 mph, a 20 degree drop in maximum temperature below that of low gear was observed.

b. Pump specifications required a minimum oil viscosity of 45 SSU. For the oil used in the system 45 SSU occurred at 200 F.

c. All convoy runs were made with the engine side panels off. When the engine side panels were in place, the result was additional drop in stabilization temperature of 24 F.

d. An additional factor which, although not measurable, and which has a direct bearing on the temperature is the nature of a convoy. Convoys have what is known as an accordion effect. At high speeds on long runs vehicles have a tendency to crowd together. When this happens, the line has to slow down and vehicles separate in order to

maintain proper spacing between them. When the fork lift truck slows down, engine speed decreases. The pump will circulate fluid at a lower rate and the oil will cool.

These factors just listed will permit safe operation in ambient temperatures up to 125 F when the temperature rise measured in the hydraulic tank with the transmission in lowest gear and engine at maximum governed speed is limited to 100 F.

While temperature could be used as the criterion to control the heat rise in the system, it is still necessary to provide the manufacturer with some guide for the selection of the components. Since it had been determined that temperature rise could be controlled by controlling the pressure drop across the main control valve, it became necessary to provide a tolerance for this factor. Analysis of the data indicated that the maximum allowable pressure which would maintain the 100 degree rise in hydraulic tank was 100 psig. This additional information would control the maximum amount of restriction which could be tolerated in the main control valve.

The limitations placed on temperature and pressure would suffice for specification requirements. However, it was still necessary to modify the equipment already issued to the field.

While there are many possible solutions to correcting the deficiency in the equipment, only four approaches were tested. These were:

a. Redesign of the hydraulic tank. As previously stated, space was at a premium, therefore little could be done in this area of increasing the reservoir volume. However, it was possible to redesign the tank so that the flow of fluid from discharge to suction would be along the tank walls thereby taking maximum advantage of surface radiation. While there was a temperature reduction, it was not of significant magnitude by itself to control temperature rise.

b. Control the flow into the directional control valve. By dividing the fluid as it was delivered from the pump it was possible to control the amount of fluid delivered to the valve. Excess fluid could be shunted directly to the reservoir tank. By controlling the amount of fluid which passed through the control valve it was possible to control back pressure.

c. Replace the hydraulic pump. By varying engine speed, it was possible to control the amount of fluid into the directional control valve. However, at a point where back pressure was reduced sufficiently and temperature stabilized within acceptable limits, engine speed was reduced to the point where other characteristics were affected materially enough to render the vehicle not operationally suitable. The effect, hydraulically, could be achieved by replacing the pump with one having smaller delivery rates.

d. Replace the directional control valve. This is the most obvious of those listed. Since the valve was the cause of heat by having internal porting such that flow was being restricted, it could be replaced with another model which had larger porting with decreased restriction.

It should be mentioned that there are various combinations of the four listed which could also offer satisfactory solutions. Each of the four solutions has disadvantages and by themselves do not offer the ideal solution. When certain combinations are made the disadvantages of each by itself are often nullified.

Since a number of vehicles had already been issued, other factors had to be taken into consideration and trade-offs had to be made before the final solution was agreed upon. Some of the factors were:

a. Economy: The pumps and directional control valves are costly units in the price range upwards of $300 each. Installation costs for directional controls would be high as it required extensive repiping.

b. Parts availability: The parts required for the item would have to be available within a reasonable time frame since operation of these vehicles affected our combat readiness.

c. Labor availability: The skills required to apply the remedy would have to be available at the level at which the actual work would be done.

d. Simplicity of installation: Consideration had to be given to the desired echelon which would apply remedy. The world wide dispersion of vehicles dictated that the remedy should be applied at the lowest possible echelon in order to avoid massive transportation costs to any single location. Therefore, if the remedy could be readily applied without the need of skilled labor, the echelon would necessarily be a low one.

e. Ease of maintenance: This item was a basic consideration. The parts required to correct the deficiency could not create an added burden on those responsible for the maintenance of the item.

Applying the factors just listed to the possible solution, it was decided that the application of a by-pass and flow divider type valve which would govern the flow from the pump to directional control valve on a priority basis would satisfy these requirements.

a. Economy: The complete installation would approximate 20% of the cost of installation of either a new pump or new directional control valve.

b. Parts availability: The item was an off-the-shelf item and readily available.

c. Labor availability: The item could be applied by an ordinary mechanic.

d. Ease of installation: The unit could be sent to the field in kit form and would require only three connections.

e. Ease of maintenance: The operation of this valve is relatively simple -- fluid from the pump goes to the inlet of the divider and through a fixed orifice in a sliding valve. The oil passing over the orifice creates a pressure differential that pushes the valve against a spring. Since the spring exerts essentially a constant force, the pressure differential is constant and the flow to create the differential is constant. If the flow through the orifice tends to increase, the pressure differential is increased and the valve moves to uncover a by-pass port. The size of the orifice may be selected for any desired amount and can be controlled within one gallon per minute. Due to the simplicity of this type valve, no additional maintenance burden would result from its installation.

Testing of this valve against the operational requirements indicated that concessions in this area had to be made as follows:

a. Engine speed had to be governed down slightly.

b. Back pressure up to 140 psig would have to be tolerated resulting in

c. A temperature rise up to approximately 143 F.

As for future procurement, the design of the hydraulic system has been changed so that a new pump and directional control valve will be used which during testing met the 100 psi back pressure and 100 F temperature rise limitations.

PRECISION OF SIMULTANEOUS MEASUREMENT PROCEDURES

W. A. Thompson, Jr.
National Bureau of Standards
Washington, D. C.

1. ABSTRACT. We consider the problem of measurement under the following conditions: The process of gathering the data is such that on any given item only one opportunity for measurement occurs, but it can be observed simultaneously by several instruments. The items to be measured are variable so that one cannot obtain replicate observations with the same instrument which would show directly the variance of the instrument readings. Procedures are discussed for estimating the precisions of the instruments and the variability of the items being measured.

An example due to Simon and Grubbs is helpful in fixing ideas. The burning times of thirty similar fuzes are determined by several different observers. We limit our discussion to the data taken by observers A and C; hence there are two determinations of the burning times of thirty different fuzes or sixty observations in all. If each of the fuzes had the same running time (which is the manufacturer's goal) and if both of the observers were absolutely accurate, then all sixty observations would be equal. However, considerable inequality in such data always occurs due to variation in the manufacturing process and inaccuracy of the observations. It then becomes desirable to use the sixty observations to answer as many questions as possible about measurement bias and precision, mean fuze running time, and variability of burning times about their mean.

2. THE MODEL. With the verbal description of the previous section in mind, consider the following mathematical formulation. Let $x_1, \ldots, x_N$ denote the true values of the items to be measured. Assume that $x_1, \ldots, x_N$ constitute a random sample of size N selected from a population having mean μ and variance σ^2. Each of the items in this sample is then measured by p instruments. y_{ij} is the measurement of the i^{th} item $(i=1, \ldots, N)$ according to the j^{th} instrument $(j = 1, \ldots, p)$. The consequence of this measurement is that an instrumentation error e_{ij}, chosen at random from the j^{th} instrument's population of errors, is added to the true value of the i^{th} item:

$$y_{ij} = x_i + e_{ij}. \tag{2.1}$$

The instrumentation errors are taken to be uncorrelated among themselves and also uncorrelated with the items selected for measurement. The mean and variance of e_{ij} are β_j and σ_j^2, respectively; β_j may be called the bias of the j^{th} instrument.

Denoting the vector $(y_{i1}, \ldots, y_{ip})$ by Y_i, we may think of $Y_1, \ldots, Y_N$ as constituting a sample of size $N = n+1$ from a p-variate distribution with mean vector $(\mu+\beta_1, \ldots, \mu+\beta_p)$ and dispersion matrix

$$\Sigma = \begin{pmatrix} \sigma^2+\sigma_1^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_2^2 & \cdots & \sigma^2 \\ \vdots & \vdots & & \vdots \\ \sigma^2 & \sigma^2 & & \sigma^2+\sigma_p^2 \end{pmatrix} . \tag{2.2}$$

Notice, in passing, that if all instrument variances are equal, then the model becomes a completely random one-way layout and may be analysed by the methods which appear, for example, in Scheffé [5].

A paragraph on notation will perhaps be helpful. ξ_j will be used as a more succinct notation for $\mu+\beta_j$, $j = 1, \ldots, p$. We will frequently write $\Sigma = (\sigma_{jj'})$ when we mean that $\sigma_{jj'}$ is the element in the j^{th} row and j'^{th} column of Σ. In the same spirit $\Sigma^{-1} = (\sigma^{jj'})$, $A = (a_{jj'})$ and $S = (s_{jj'})$ will be common notations. Here

$$a_{jj'} = \sum_{i=1}^{N} (y_{ij} - \bar{y}_{\cdot j})(y_{ij'} - \bar{y}_{\cdot j'}) , \tag{2.3}$$

and $s_{jj'}$ is the usual unbiased estimate of $\sigma_{jj'}$, i.e. $s_{jj'} = a_{jj'}/n$.

3. <u>POINT ESTIMATES.</u> In [2], Grubbs recommends certain estimates of item and instrument variance. For p = 2 instruments, these estimates are

(3.1) $\sigma^2 \sim s_{12}$, $\sigma_1^2 \sim s_{11} - s_{12}$ and $\sigma_2^2 \sim s_{22} - s_{12}$,

where $\sim$ is to be read "is estimated by". For $p \geq 3$, Grubbs recommends

$$\sigma^2 \sim \frac{2}{p(p-1)} \sum_{j<j'} s_{jj'},$$

$$(3.2) \qquad \sigma_1^2 \sim s_{11} - \frac{2}{p-1} \sum_{j=2}^{p} s_{1j} + \frac{2}{(p-1)(p-2)} \sum_{2 \leq j<j'} s_{jj'},$$

with an analogeous estimate of the other instrument variances. Gaylor [1] shows that Grubbs' estimate of σ^2 is equivalent to a familiar variance component estimate. These estimates are reasonable in that they have the correct dimensionality, are unbiased, and have appropriate symmetry properties when the instrument labels are interchanged. Further, if the underlying distribution is normal, then Grubbs' estimates are simple functions of the sufficient statistics and in the case $p=2$ they have a maximum likelihood property. However, as Grubbs has verbally pointed out his estimates are unreasonable in that they frequently assume negative values even though the parameters themselves must be non-negative by their very definition.

For $p=2$ this objectional characteristic has been eliminated in [8]; here the altered estimates of table 1 have been proposed. The top line of this table yields the estimates (3.1) under conditions where they are non-negative. The remaining entries show how Grubbs' estimates can be modified when negativity would result from using (3.1). These modified estimates have been derived, under normality assumptions, from a principle of restricted maximum likelihood which is fairly well accepted in other brances of statistical practice. A tilda placed over a parameter indicates its restricted maximum likelihood estimate.

Table 1. Non-negative estimates in the two instrument-case.

Conditions	$\tilde{\sigma}^2$	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$
$s_{11} \geq s_{12}$ $s_{22} \geq s_{12} \geq 0$	s_{12}	$s_{11} - s_{12}$	$s_{22} - s_{12}$
$s_{22} \geq s_{12} > s_{11}$	s_{11}	0	$s_{11} + s_{22} - 2s_{12}$
$s_{11} \geq s_{12} > s_{22}$	s_{22}	$s_{11} + s_{22} - 2s_{12}$	0
$s_{12} < 0$	0	s_{11}	s_{22}

4. RELATIVE PRECISION. It is clear that if the instrumentation of an experiment is to be effective then the instruments must be precise relative to the variability of the quantity being measured. A frequently quoted rule of thumb is that the instrument precision should be an order of magnitude greater than that of the item being measured. Such a statement has no firm meaning unless a measure of instrument precision and a measure of item variability have been agreed upon. Here we adopt $\Delta_1 = \sigma/\sigma_1$ as a measure of the relative precision of the first instrument. Then, for example, the above mentioned rule of thumb would become $\Delta_1 \geq 10$.

In the two-instrument case, assuming normality, we may use a result of Roy and Bose [3] to make inferential statements of a statistical nature concerning the parameter Δ_1. In our terminology their result states that

$$(4.1) \qquad a_{11}\left[\frac{n-1}{|A|}\right]^{\frac{1}{2}}\left(\frac{a_{12}}{a_{11}} - \frac{\sigma_{12}}{\sigma_{11}}\right)$$

has the t-distribution with n - 1 d.f. where $|A| = a_{11}a_{22} - a_{12}^2$. Noting that $\sigma_{12}/\sigma_{11} = (1+\Delta_1^2)^{-1}$, we may verify that the quantity (4.1) is less than t_α, if and only if

$$(4.2) \qquad \Delta_1^2 > \frac{a_{12} - t_\alpha\left(\frac{|A|}{n-1}\right)^{\frac{1}{2}}}{a_{11} - a_{12} + t_\alpha\left(\frac{|A|}{n-1}\right)^{1/2}}.$$

Hence if t_α is the upper α percentage point of the t- distribution with n-1 d.f. then the square root of the right hand side of (4.2) provides a lower confidence bound for Δ_1, the confidence coefficient being $1-\alpha$. The inequality (4.2) can also be used for the purpose of hypothesis testing For example, we may reject $\Delta_1 \geq 10$ at the significance level α if (4.2) is violated with $\Delta_1^2 = 100$.

5. A SIMULTANEOUS CONFIDENCE REGION. For some purposes it may not be enough to consider relative precision; we may be interested in the actual non-relative precisions and the item variability. Estimation of

the parameters $\sigma^2, \sigma_1^2, \ldots, \sigma_p^2$ was discussed in section 3; but how reliable are estimates? This question is dealth with in [9], again under assumptions of normality.

In the two-instrument case, the probability is at least $1 - 2\alpha$ that the following three relations hold simultaneously

$$
(5.1) \qquad
\begin{aligned}
&\left| \sigma^2 - a_{12} K \right| \leq M(a_{11}a_{22})^{\frac{1}{2}} , \\
&\left| \sigma_i^2 - (a_{ii} - a_{12})K \right| \leq M \left[a_{ii}(a_{11} + a_{22} - 2a_{12}) \right]^{\frac{1}{2}}; \quad i = 1, 2.
\end{aligned}
$$

Here K and M are to be found in Table 2 under the desired value of 2α.

Table 2. The table gives values of K and M which yield 1 - 2α confidence regions when used in conjunction with the relations (5.1).

2α	.01		.05	
n	K	M	K	M
3	99.78	99.72	19.79	19.71
4	12.38	12.33	4.146	4.077
5	3.980	3.931	1.726	1.665
6	1.903	1.353	.9636	.9088
7	1.120	1.078	.6290	.5786
8	0.7459	.7076	.4516	.4032
9	0.5389	.5031	.3453	.3022
10	0.4120	.3782	.2761	.2357
11	0.3282	.2963	.2280	.1901
12	0.2698	.2395	.1932	.1573
13	0.2272	.1983	.1663	.1328
14	0.1951	.1675	.1464	.1140
15	0.1702	.1438	.1301	.09925
16	0.1505	.1251	.1169	.08738
17	0.1344	.1100	.1060	.07767
18	0.1213	.09772	.09632	.06962
19	0.1103	.08752	.08904	.06287
20	0.1009	.07896	.08237	.05713
22	.08610	.06546	.07152	.04795
24	.07484	.05538	.06311	.04098
26	.06605	.04763	.05641	.03554
28	.05901	.04152	.05096	.03121
30	.05328	.03660	.04644	.02768
35	.04272	.02778	.03796	.02127
40	.03556	.02200	.03205	.01700
45	.03040	.01797	.02771	.01398
50	.02652	.01503	.02440	.01176
60	.02109	.01110	.01967	.00875
70	.01748	.00862	.01646	.00684
80	.01492	.00694	.01415	.00553
90	.01300	.00575	.01241	.00460
100	.01152	.00486	.01104	.00390

Page Blank

For more than two instruments a similar result is valid. We have the following relations with probability at least $1 - 2\alpha$

$$(5.2) \qquad \max_{j \neq j'} \left[a_{jj'} K - M (a_{jj}\, a_{j'j'})^{\frac{1}{2}} \right]$$

$$\leq \sigma^2 \leq \min_{j \neq j'} \left[a_{jj'} K + M (a_{jj}\, a_{j'j'})^{\frac{1}{2}} \right],$$

$$\max_{j \neq 1} \left\{ (a_{11} - a_{1j}) K - M \left[a_{11} + a_{jj} - 2a_{1j} \right]^{\frac{1}{2}} \right\}$$

$$\leq \sigma_1^2 \leq \min_{j \neq 1} \left\{ (a_{11} - a_{1j}) K + M \left[a_{11} + a_{jj} - 2a_{1j} \right]^{\frac{1}{2}} \right\},$$

plus $p - 1$ similar inequalities involving $\sigma_2^2, \ldots, \sigma_p^2$. Unfortunately, for p in excess of two, tables of K and M are unavailable. The only result which is currently ready for use is an approximation valid for large n: Choose ℓ to satisfy $P(\ell \leq \chi^2_{n-p+2}) = 1 - 2\alpha$, write $K = M = 1/2\ell$ and substitute this common value in (5.2). I feel obliged to point out that for $p = 2$, the only case where exact values are available, this approximation is rather poor.

6. NUMERICAL EXAMPLE. Returning to the fuze burning time data, we may identify observer A as the first instrument and observer C as the second. From Table I of Grubbs' paper [2] we obtain $a_{11} = 1.3671$, $a_{22} = 1.3227$, $a_{12} = 1.3320$ and $n = 29$. From the third row entry of our table 1, we estimate $\tilde{\sigma} = .21$, $\tilde{\sigma}_1 = .03$ and $\tilde{\sigma}_2 = 0$. By the method of section 4 we obtain, for example, that the relative precision of observer C exceeds 5.1 with a confidence of 95%. Alternatively, from a hypothesis testing point of view we would reject the rule of thumb requirement, $\Delta_2 \geq 10$, at the 5% level. The relations (5.1) and table 2 yield the following 95% simultaneous confidence region: $.16 \leq \sigma \leq .32$, $0 \leq \sigma_1 \leq .09$ and $0 \leq \sigma_2 \leq .07$. In calculating these simultaneous confidence intervals we have replaced all negative lower bounds by zero. Notice that the confidence intervals bracket their respective estimates and hence, in the confidence region sense, indicate the uncertainty of these estimates.

I would like to express my appreciation to the National Academy of Sciences-National Research Council for granting me the opportunity to participate in their postdoctoral research program, and in particular to the Natiönl Bureau of Standards for its support of my research under this program. Also I wish to thank Mrs. Marion Carson for her help in carrying out the computations for table 2.

REFERENCES

[1] D. W. Gaylor, "Equivalence of two estimates of product varinace", J. Amer, Stat, Assoc., vol. 51 (1956), pp. 451-453

[2] Frank E. Grubbs, "On estimating precision of measuring instruments and product variability", J. Amer. Stat. Assoc., vol 43 (1948), pp. 243-264.

[3] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," Ann. Math. Stat., vol. 24 (1953), pp. 513-536.

[4] Thomas S. Russel and Ralph Allan Bradley, "One-way variances in a two-way classification", Biometrika, vol. 45 (1958), pp. 111-129

[5] Henry Scheffe, The Analysis of Variance, John Wiley and Sons, Inc., N. Y., 1959.

[6] L. E. Simon, "On the relation of instrumentation to quality control", Instruments, vol. 19 (1946), pp. 654-656.

[7] H. Fairfield Smith, "Estimating precision of measuring instruments", J. Amer. Stat. Assoc., vol. 45 (1950), pp. 447-451.

[8] W. A. Thompson, Jr., "The problem of negative estimates of variance components", Ann. Math. Stat., vol 33 (1962), pp. 273-289.

[9] W. A. Thompson, Jr., "Estimation of dispersion parameters", in preparation.

APPLIED MICROSCOPY

George I. Lavin
Terminal Ballistics Laboratory
USA Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

The Aberdeen Proving Ground in general and the Ballistic Research Laboratories in particular are interested in the manner by which absorbed energy is utilized by various systems. The technique describes the possibility of an evaluation of the traumatic effect brought about by the absorption of energy at the cellular level.

This paper (presented, in brief, to the Eighth Conference on the Design of Experiments in Army Research, Development and Testing, The Walter Reed Army Institute of Research, Washington, D. C., October 1962) has to do with the ultraviolet microscopy of biological tissues.

There is much controversy having to do with the structure and function of biological cellular configurations. This controversy will continue to exist until methods are employed for the analysis and evaluation of experimental results which do not (by preparation and/or examination) change the materials during the evaluation procedure.

A paper describing a photomicrographic technique was published by Köhler [1] in 1904. The optical system was quartz instead of the usual glass. The purpose of this was to take advantage of the increase in resolution which is realized when ultraviolet light is used as the source of illumination instead of the visible region of the spectrum.

The light source employed by Köhler was a condensed electrical spark between metallic electrodes. This is unsatisfactory for the illumination of a microscope. The electrodes burn away non-uniformally which means that it is difficult to keep an image of the source centered on the microscope condenser. Also the noise of the spark is objectionable. In spite of these and other difficulties, some work has been published [2].

The early workers were attracted to this field because the technique offered the possibility of considerable ultimate magnification (since the resolution was about twice that obtained with visible light) although the initial magnification might be low.

In other words, the increase in resolution made possible the visualization of microscopic structure which was not evident until it was revealed by the enlarging camera.

It is to be pointed out that the obvious lack of light intensity was due to the inefficiency of the monochromator system used to isolate the wave length of light (2750A) for which the objectives were corrected.

It was also realized that tissue components such as proteins, nucleic acids and nucleoproteins have specific absorption bands in this region of the spectrum so that it should be possible to obtain micrographs of fresh and unstained material.

Thus, in a sense of the word, it can be said that the chemical constituents of the tissue can act as their own specific light absorbing medium -- the stain.

It is well understood that the fixation process (necessary for the manipulation of tissue cutting) may contribute to the micrographical picture. We do not always know what contribution this represents to our over-all picture. By means of frozen sections we can minimize this uncertainty. We will also need the frozen section-type of section for our analytical procedures.

The inherent objection to previous procedures was that it was not possible to obtain sufficient light intensity of the proper wave length (in the plane of the photographic plate) so that a field and focus could be localized and imaged with the wave length of light which is specifically absorbed by the material under examination.

In some cases, a fluorescent plate was placed directly over the microscope eyepiece. Thus another adjustment had to be made in order to bring the image in focus in the same plane of the photographic plate. In other words, the focus must be in the plane of the photographic plate.

In 1943 a short paper on the subject of ultraviolet microscopy was published by Lavin(3). A procedure was described whereby an image of the material on the stage of a microscope could be visualized by a fluorescent screen which was temporarily placed in the position usually occupied by a photographic plate. That is, in the plate holder.

The 2537 Angstrom mercury line was the light source (loc. cit.). It has been previously pointed out that this wave length of light is in the region of absorption of those chemical compounds which are to be associated with tissue structure.

The photographs which are now shown will serve to illustrate this technique and some of the results obtained by application of this procedure. The original plates were taken at a magnification of about 200 diameters.

DESCRIPTION OF PLATES

1. Ultraviolet photomicrographic apparatus described by Köhler. Note the spark source, the monochromator and the eyepiece focusing attachment.
2. Apparatus used in the present work. Light source, liquid filter, willemite focusing screen.
3. Light source -- quartz resonance radiation lamp.
4. Spectrum of light source, with and without filter.
5. Fresh Hamster muscle -- teased out (not cut), fresh, unstained.
6. Muscular dystrophy -- cross section.
7. Muscular dystrophy -- longitudinal.
8. Fresh smear of cells from a chicken egg.
9. Liver, normal -- fixed, unstained.
10. Liver, infectious hepatitis, fixed, unstained.
11. Cross section of a plant root (Sorrell), unstained.
12. Cross section, skin (mal del pinta), unstained.
13. Salivary gland of a mosquito, fresh -- unfixed, unstained.
14. Enlargement of a portion of the salivary gland shown above.

15. Kidney, cross section -- unstained.

16. Red cells, smear, monkey, dried.

17. Red cells, smear, chicken, dried, showing nuclei.

18. Arbacia eggs, showing the "relayering" on rupture.

19. Arbacia eggs, showing the results of photography in the visible, ultraviolet and infrared regions of the spectrum.

20. Muscle photographed using the 2537 A^o mercury line, unstained, visible -- stained, desicated.

21. Absorption spectrum curve of 1, 2 -- Benzanthracene.

22. Absorption spectrum of the same substance using the continuous light from a hydrogen discharge tube as the light source.

23. Apparatus for photomicroscopy -- three light sources, infrared, visible, ultraviolet.

THE USE OF CONTINUUM AS A LIGHT SOURCE FOR ABSORPTION SPECTRA. As has been pointed out those compounds such as proteins nucleic acids, nucleoproteins etc. have broad absorption bands when measured by spectrophotometers.

It has also been shown by Lavin and Northrop[4] and others that a considerable amount of band structure can be demonstrated (at room temperature) by the use of a spectrograph of comparative low dispersion. It was also indicated that these bands can be interpreted in terms of the component parts of the molecule by Lavin, Loring and Stanley[5]. This technique has also been applied to complicated mixtures (body fluids) by Dobriner, Lavin, and Rhoads[6 & 7].

Photographs illustrating the above are shown in Plate 22. It is thought that efforts to apply this technique in obtaining the absorption spectra of materials on the stage of a microscope might be worthwhile. This could be a clue to the chemical composition of the various sections of material under examination.

SUMMARY. Photomicrographs of fresh and of unstained tissues obtained with the 2537 A° mercury line as the light source are shown and some of the implications of this technique are discussed.

The possibility of a microspectrographic application to the problem is considered.

ACKNOWLEDGEMENT. The photomicrographs of tissue which were made while the author was at the Rockefeller Institute for Medical Research are used with the kind permission of Dr. Detlev W. Bronk.

REFERENCES TO PREVIOUS WORK

(1) A. Kohler, Zeits. wiss, Mik. 21, 273 (1904).

(2) F. F. Lucas, J. Franklin Inst. 217, 661 (1934).
R. W. G. Wyckoff, Cold Spring Harbor Symposia 2, 39 (1936).
J. E. Barnard, J. Roy. Microscop. Soc. 56, 365 (1936).
T. Caspersson, Skand. Arch. Physiol. 73, Suppl. 8 (1936).
J. S. Mitchell, Brit. J. Exptl. Path. 23, 285 (1942).

(3) G. I. Lavin, Rev. Sci. Instr. 14, 375 (1943).

(4) G. I. Lavin and J. H. Northrop, J. Am. Chem. Soc. 57, 874 (1935).

(5) G. I. Lavin, H. S. Loring, and W. M. Stanley, J. Biol. Chem. 130, 259 (1940).

(6) K. Dobriner, G. I. Lavin, and C. P. Rhoads, Cancer Research 2, 79 (1942).

(7) K. Dobriner, C. P. Rhoads, and G. I. Lavin, Cancer Research 2, 95 (1942).

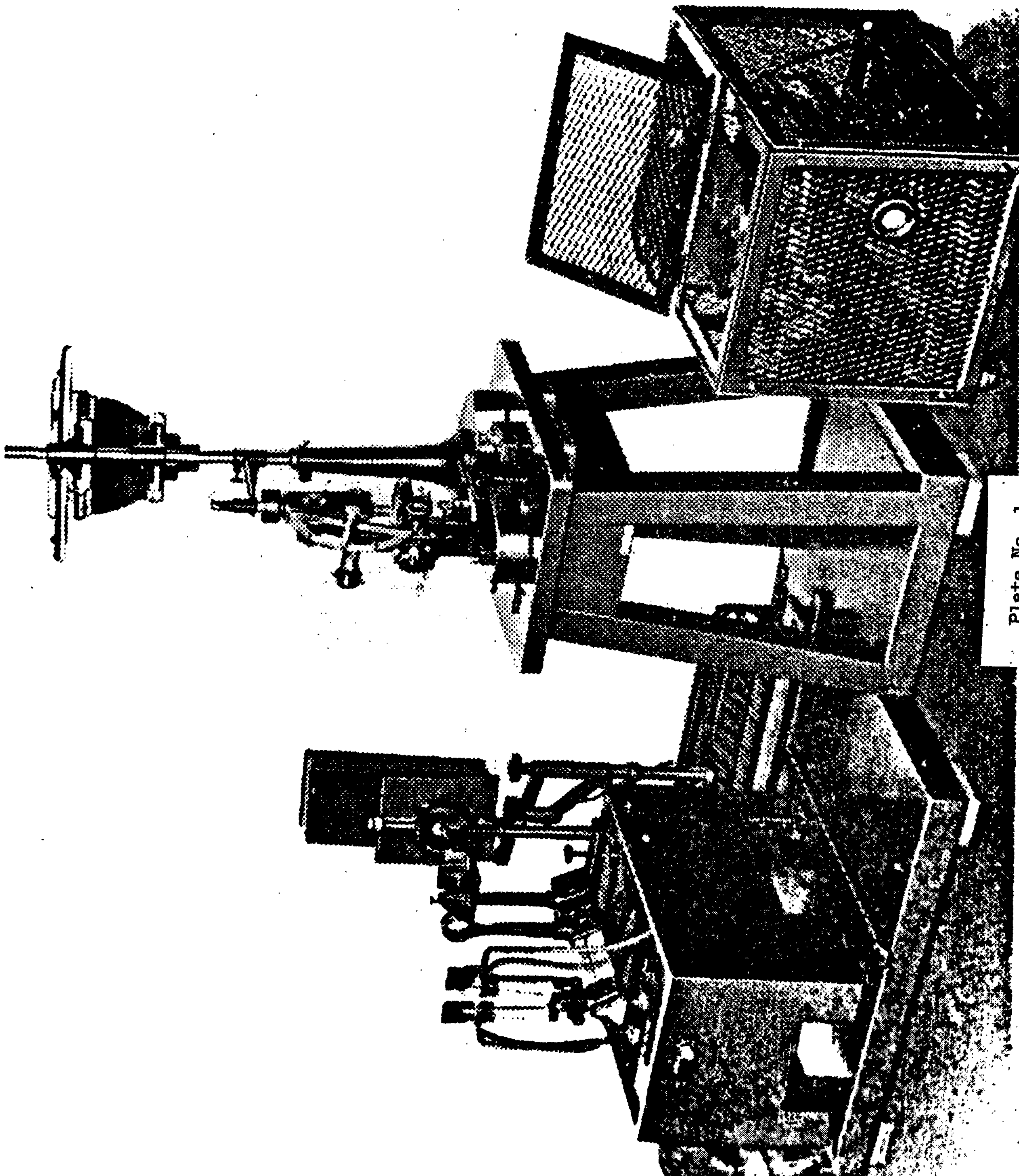


Plate No. 1

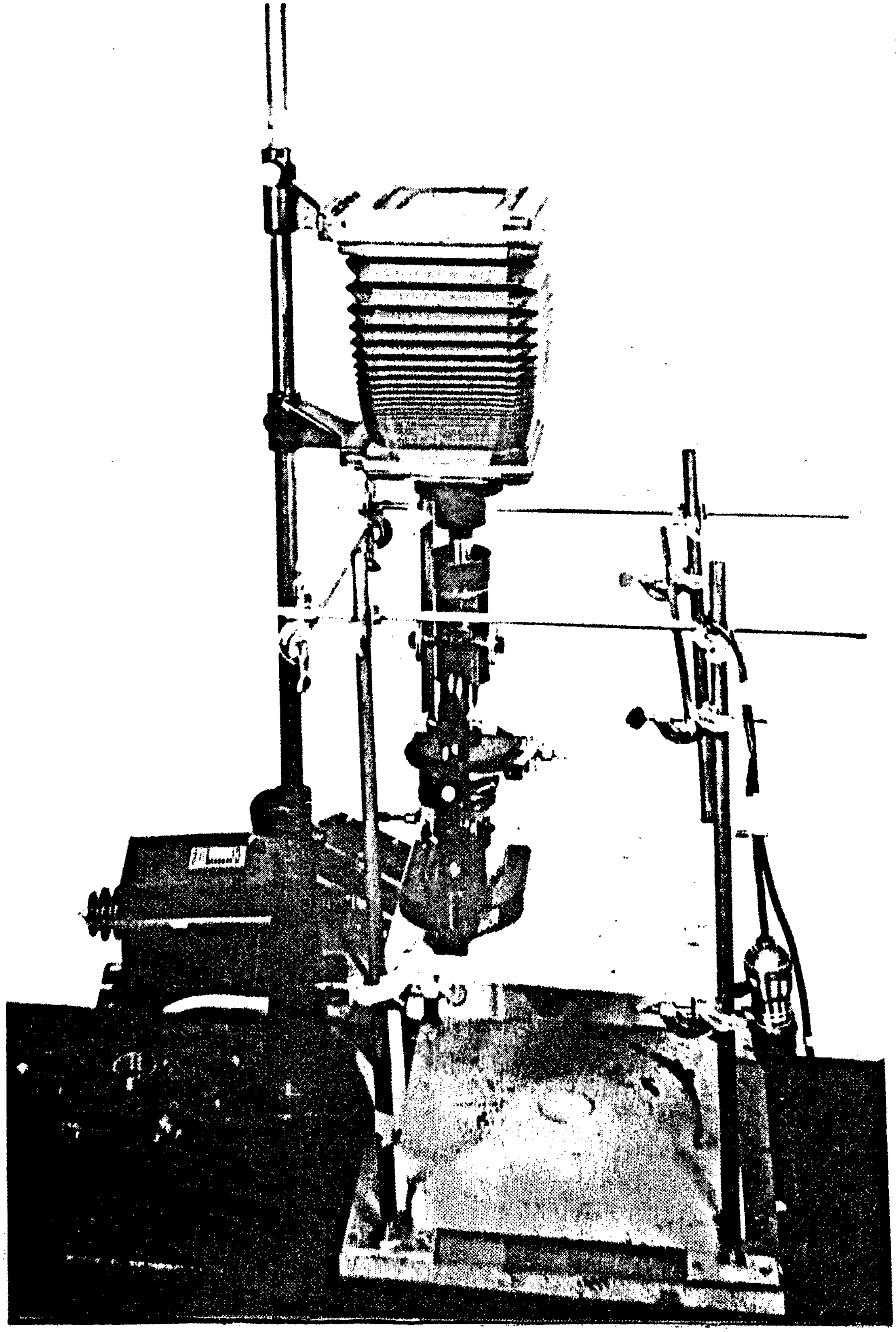


Plate No. 2

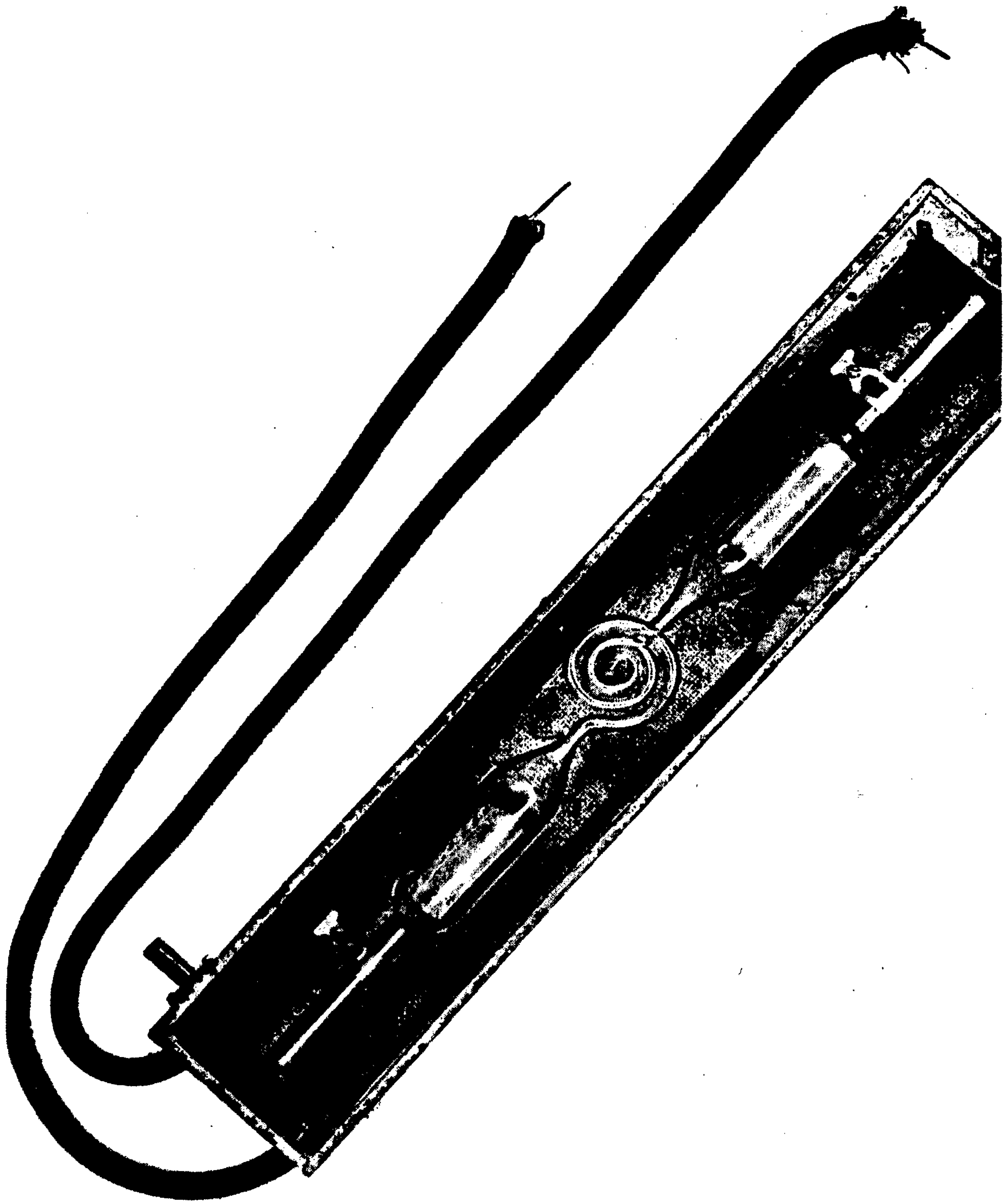


Plate No. 3

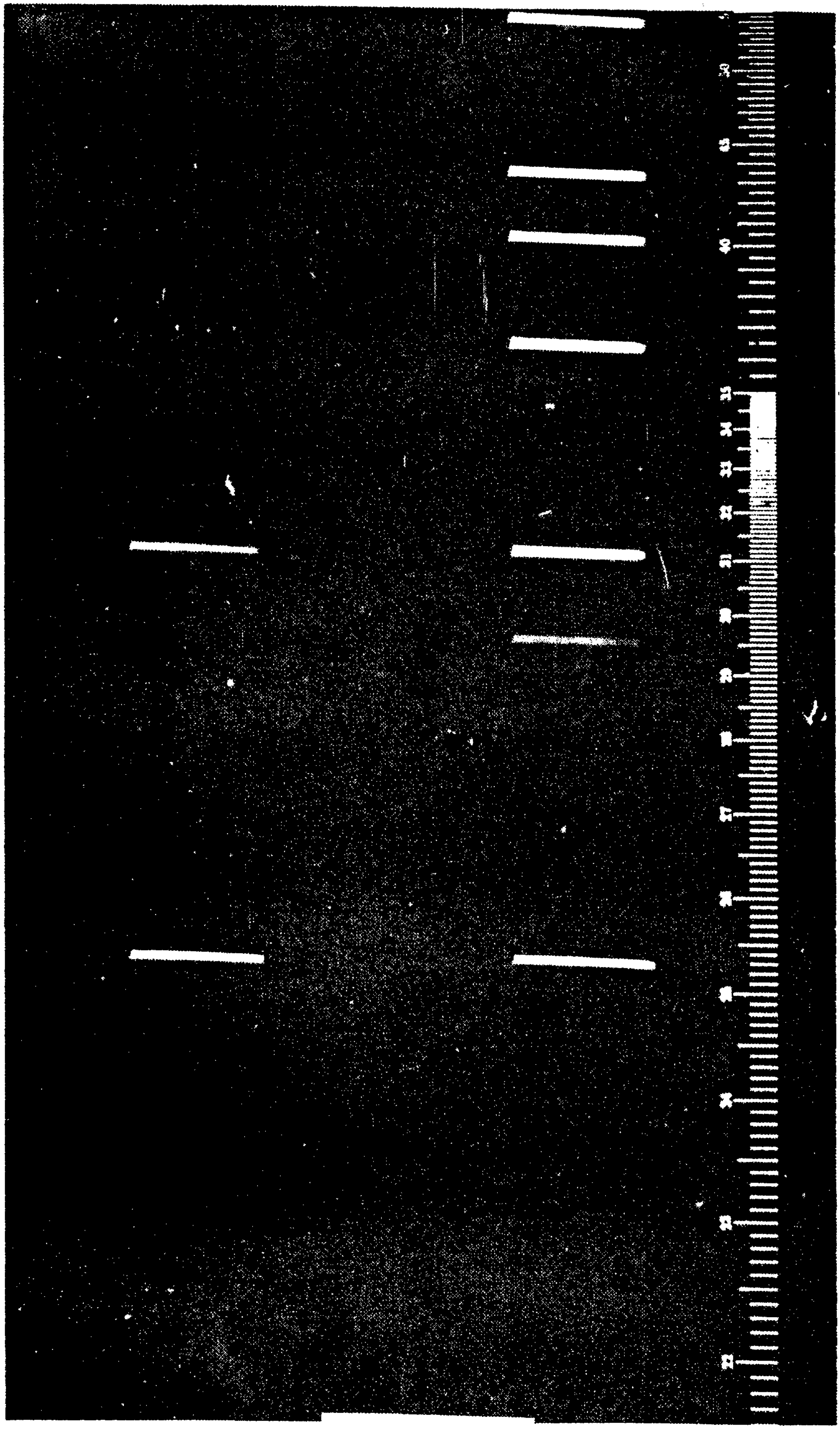


Plate No. 4

Plate No. 5

Plate No. 6

Plate No. 7

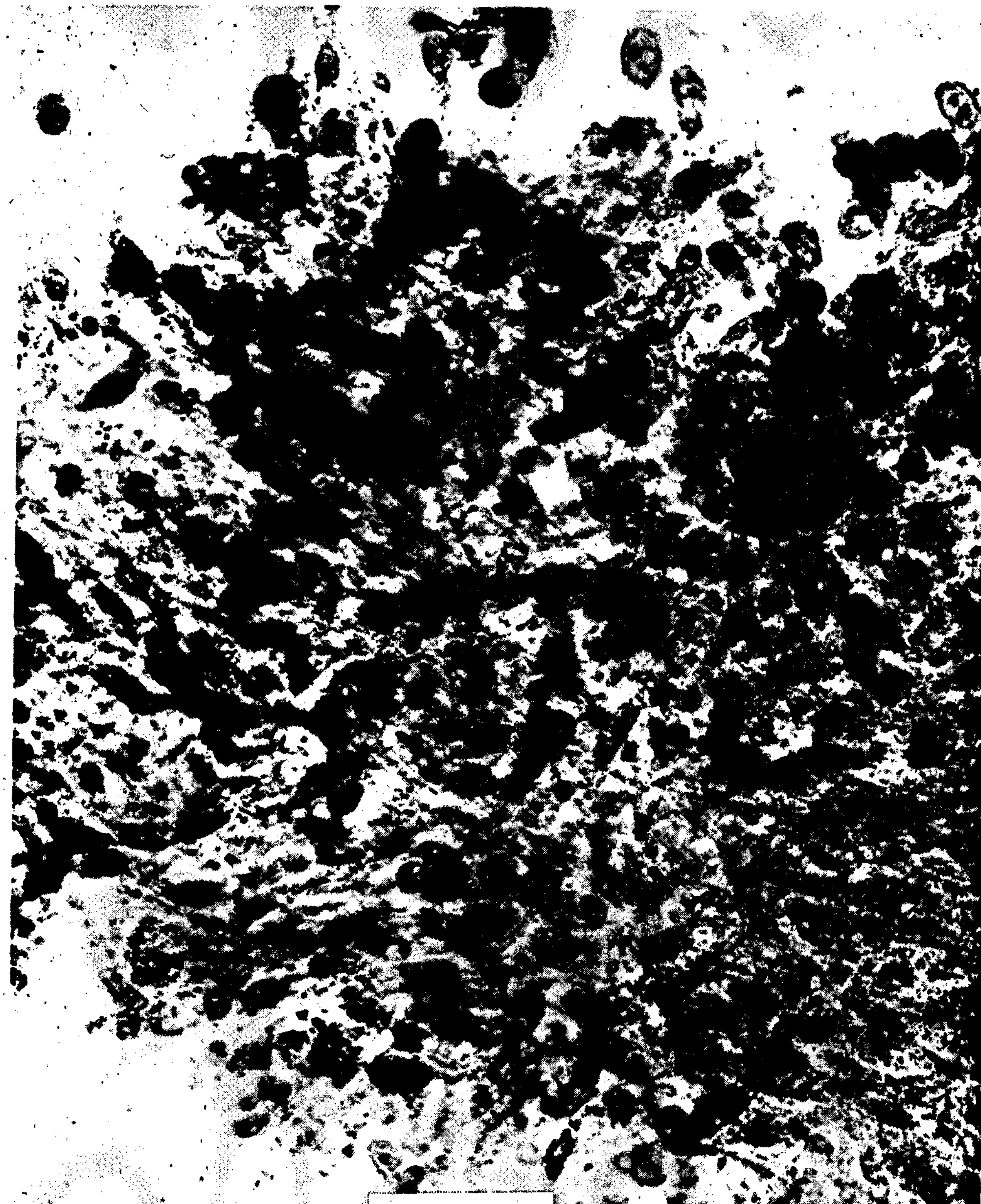


Plate No. 8

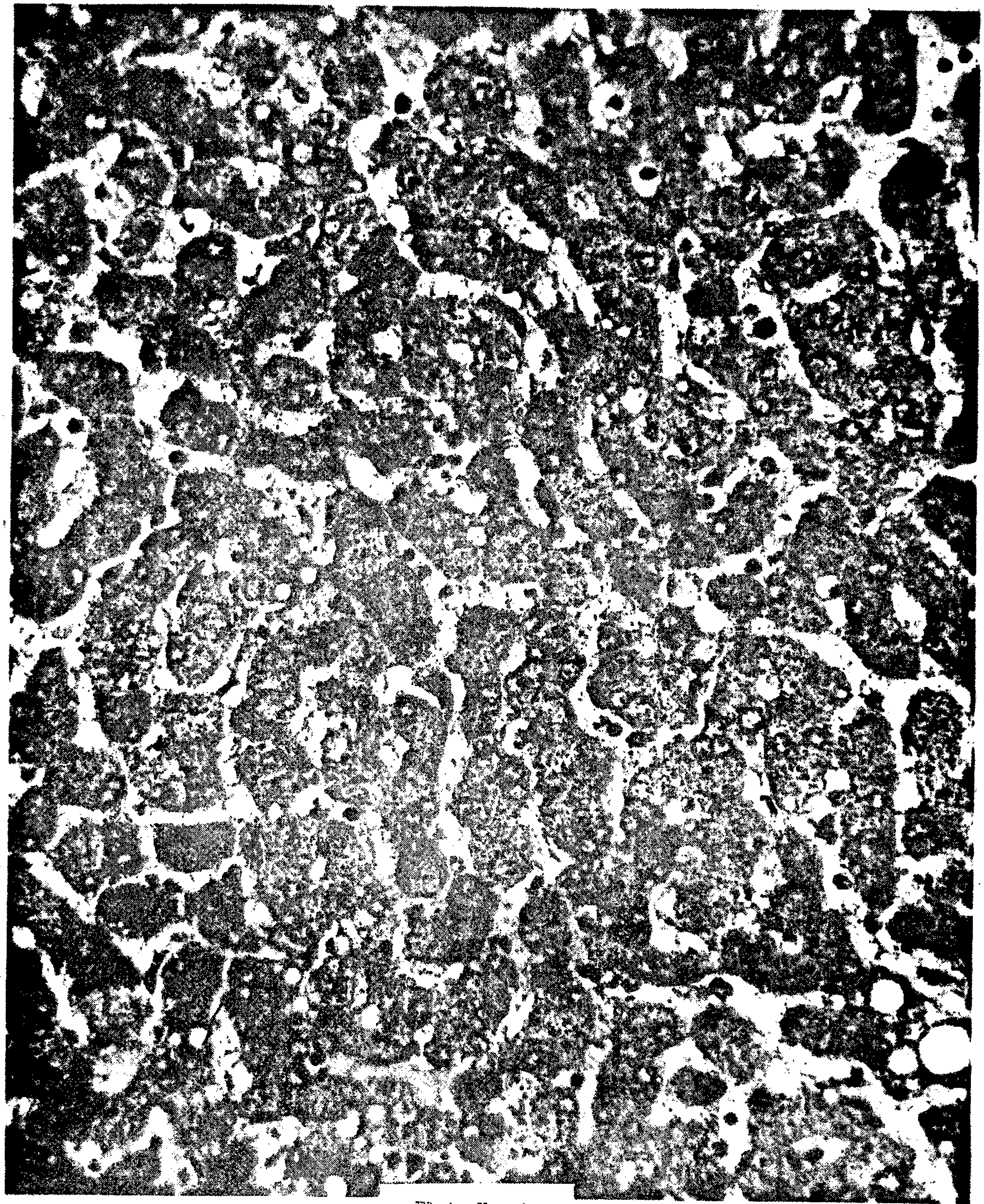


Plate No. 9

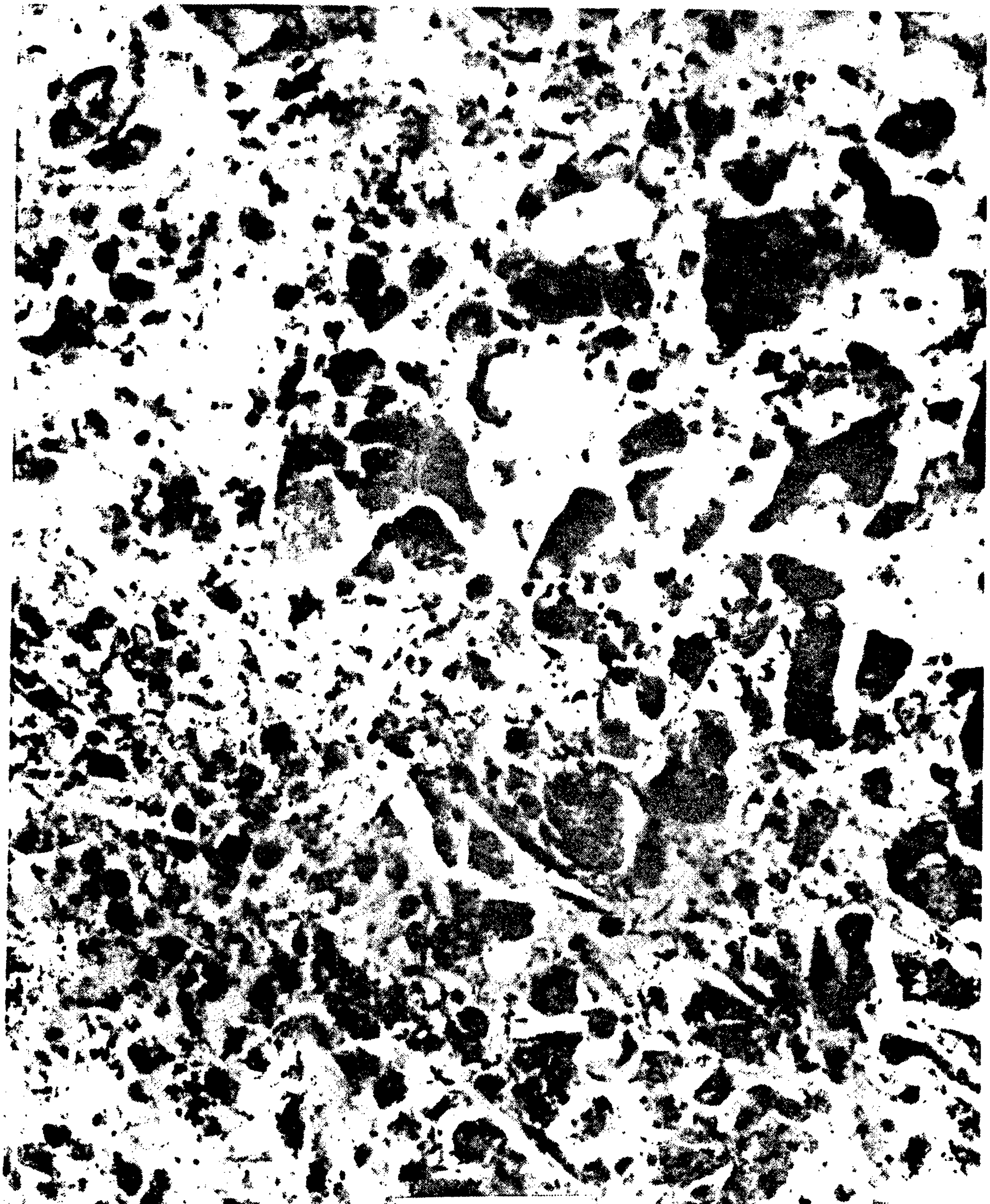


Plate No. 10

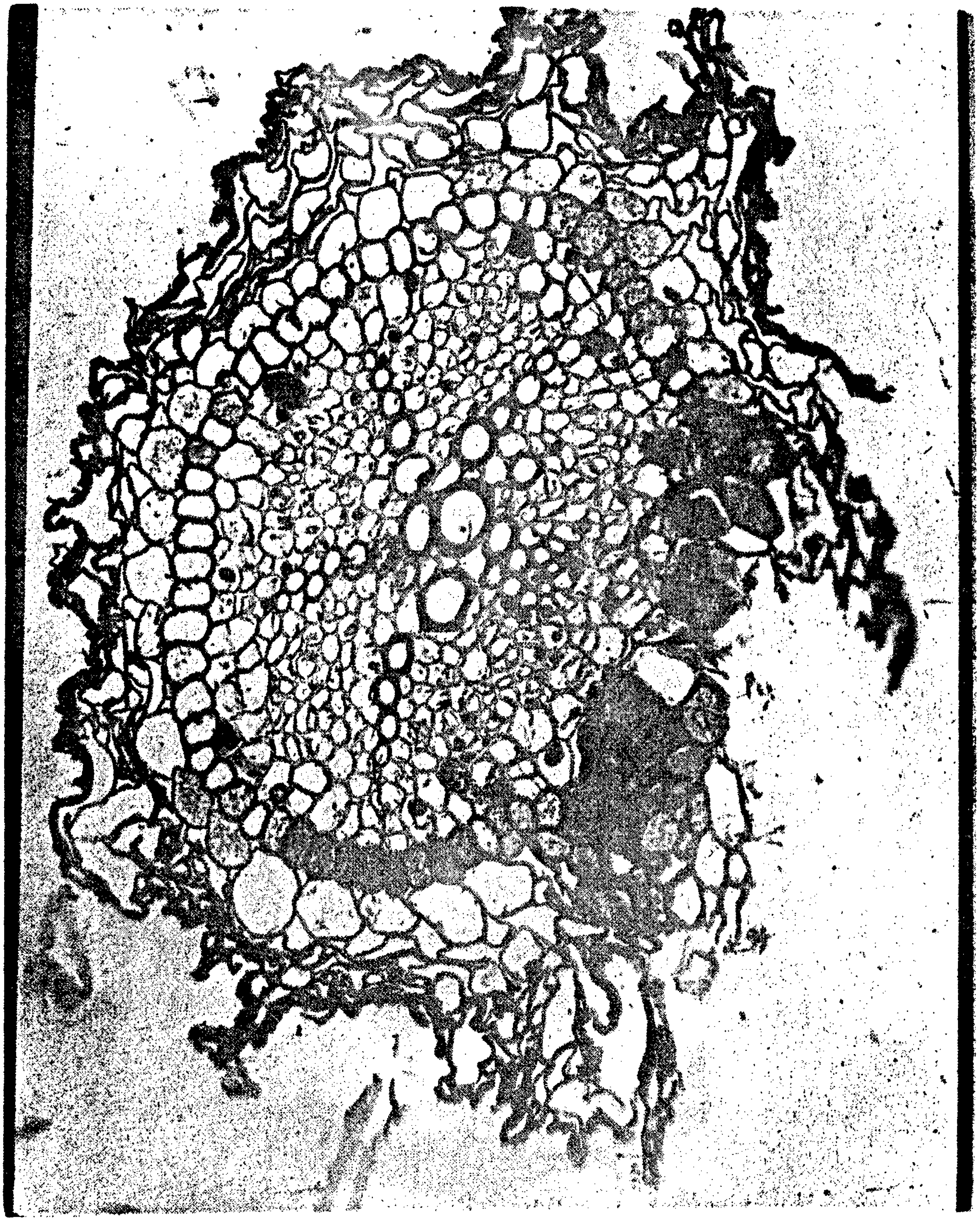


Plate No. 11

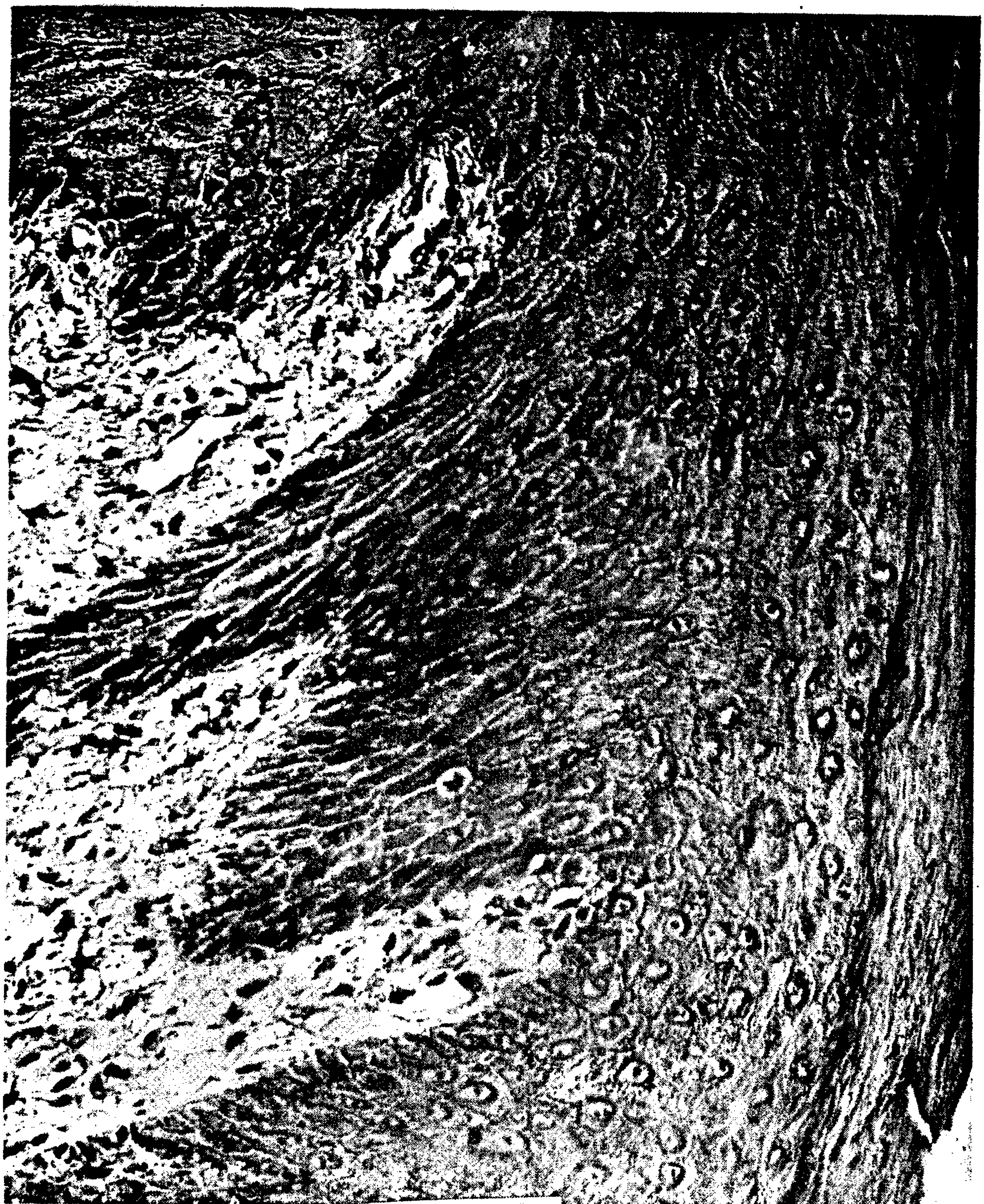


Plate No. 12

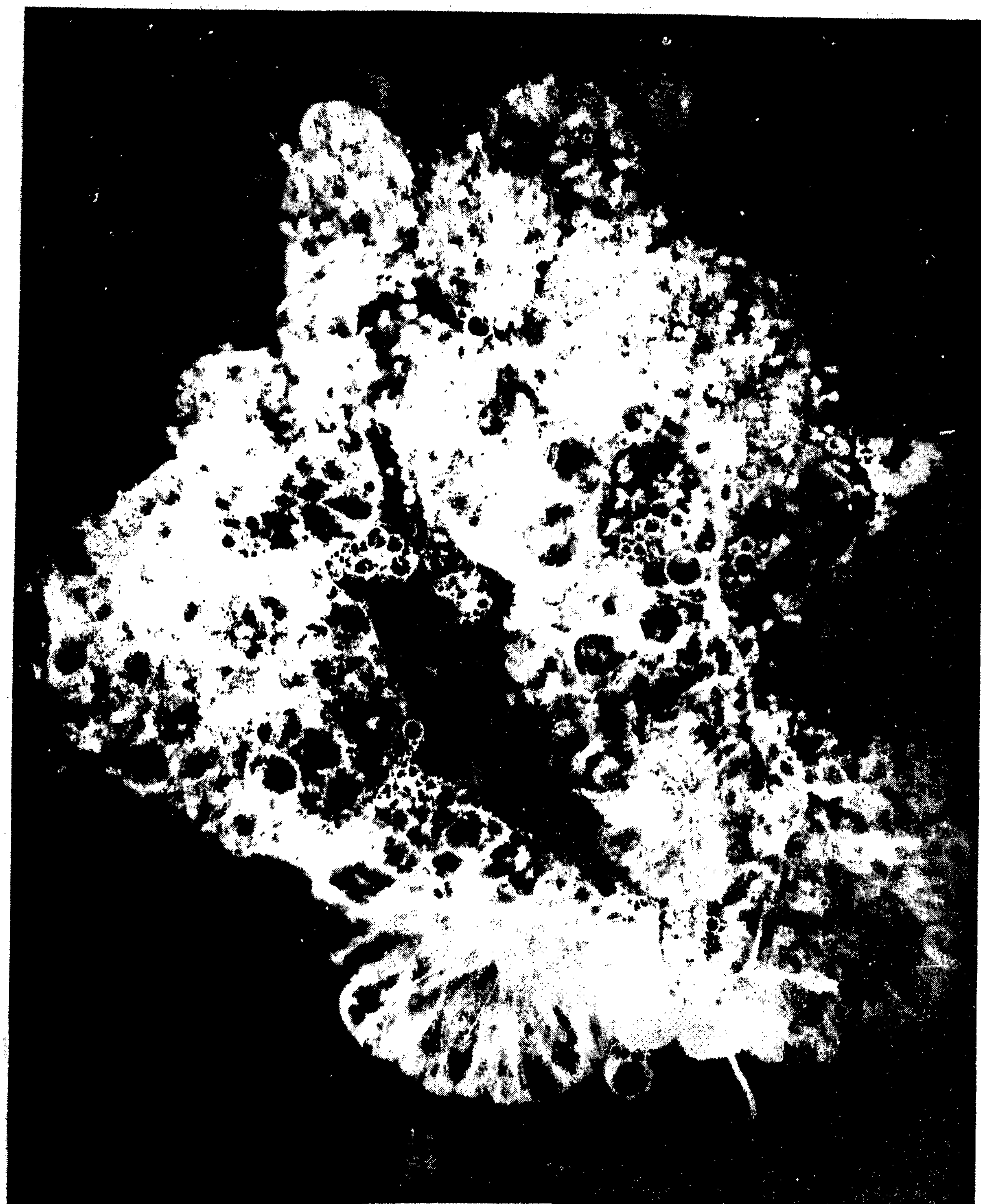


Plate No. 13

Plate No. 14

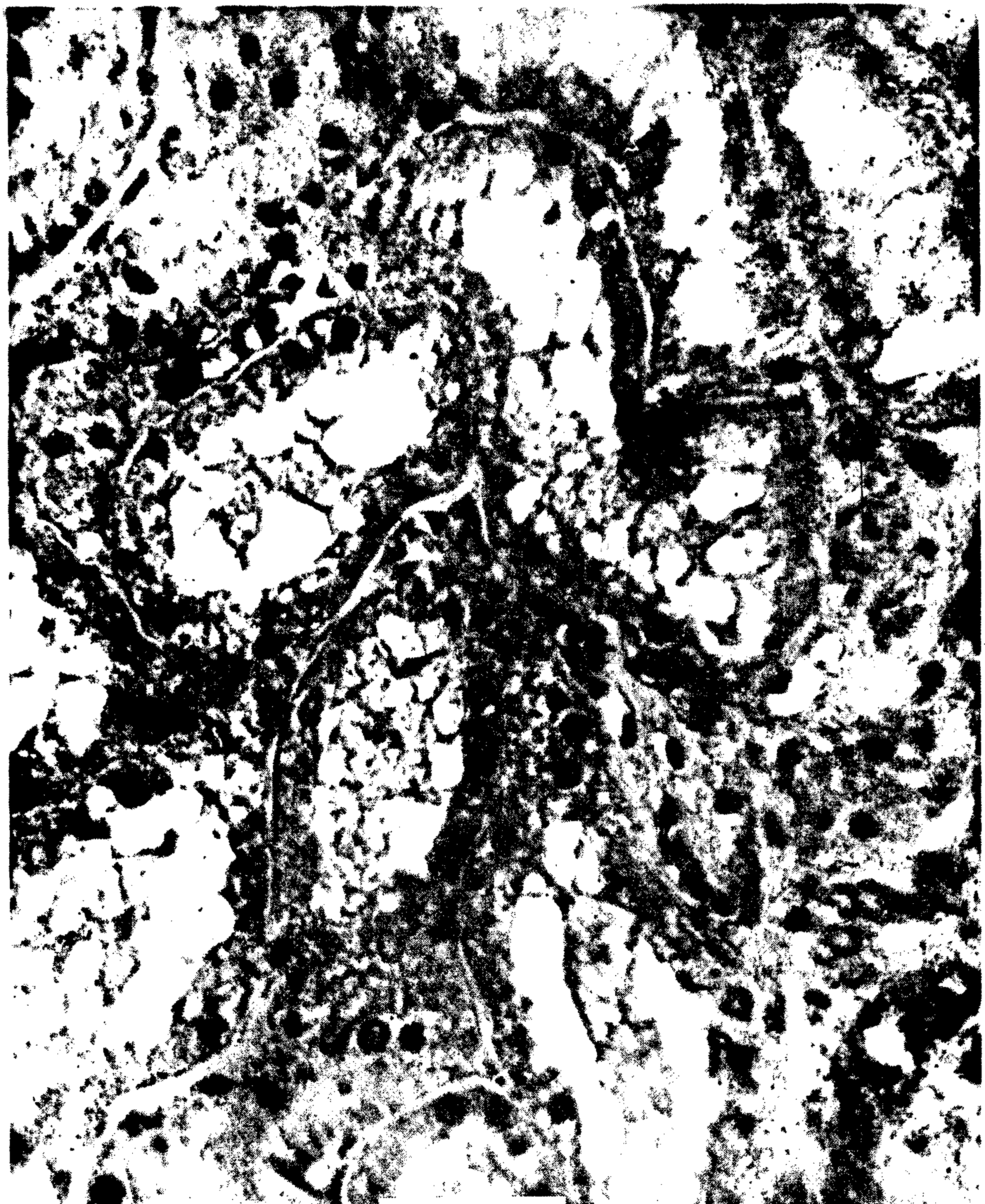


Plate No. 15

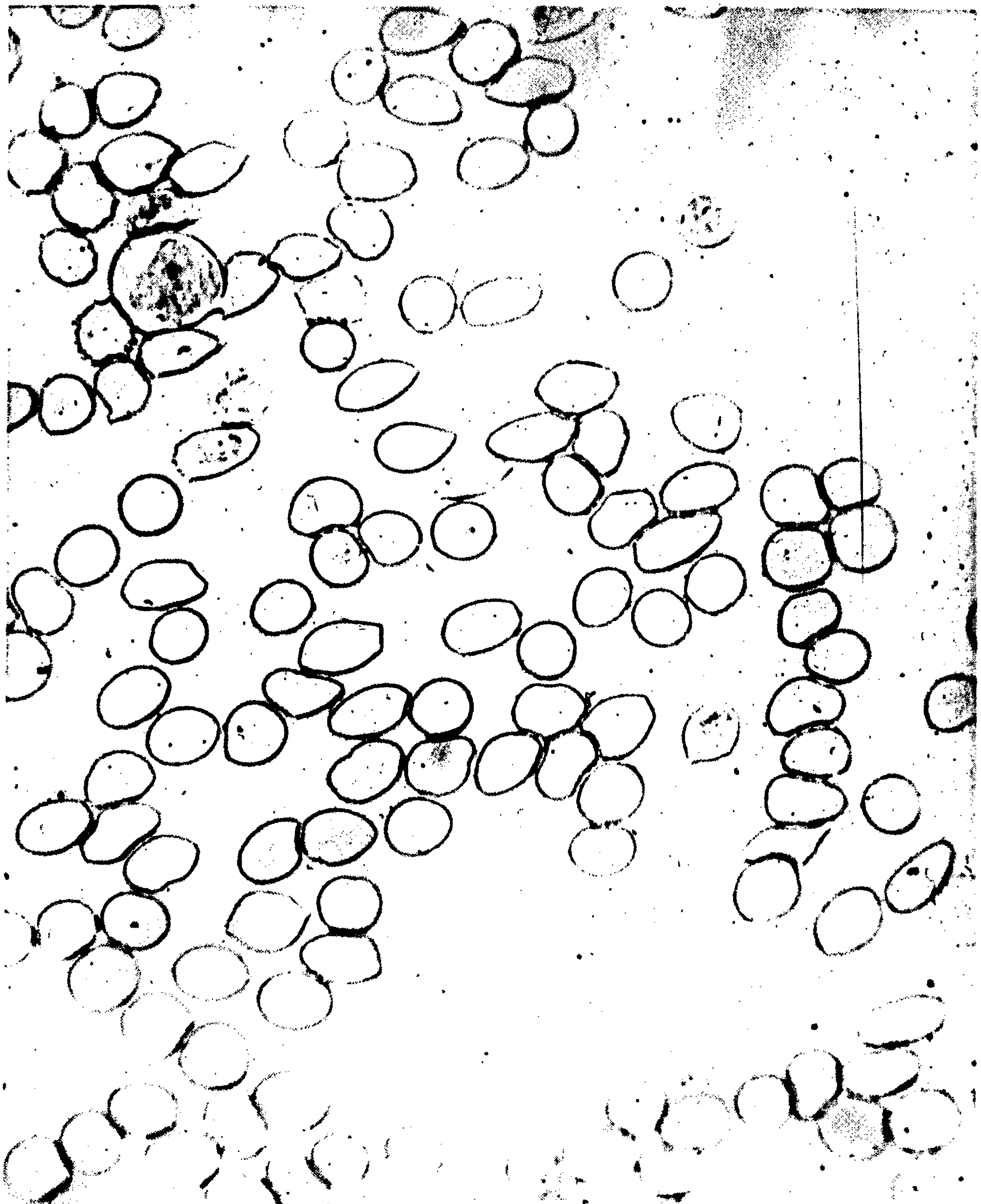


Plate No. 16

Plate No. 17

Plate No. 18

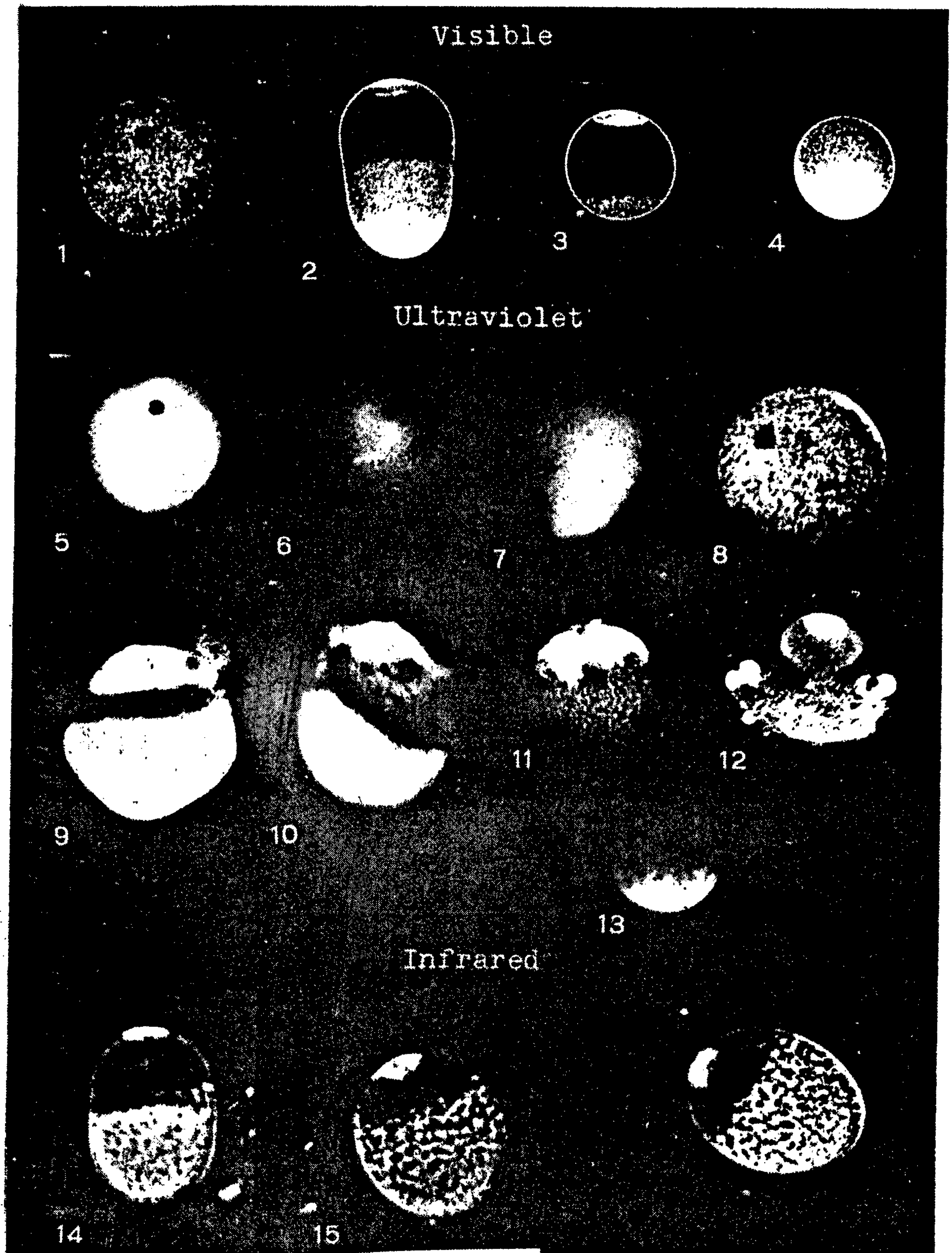


Plate No. 19

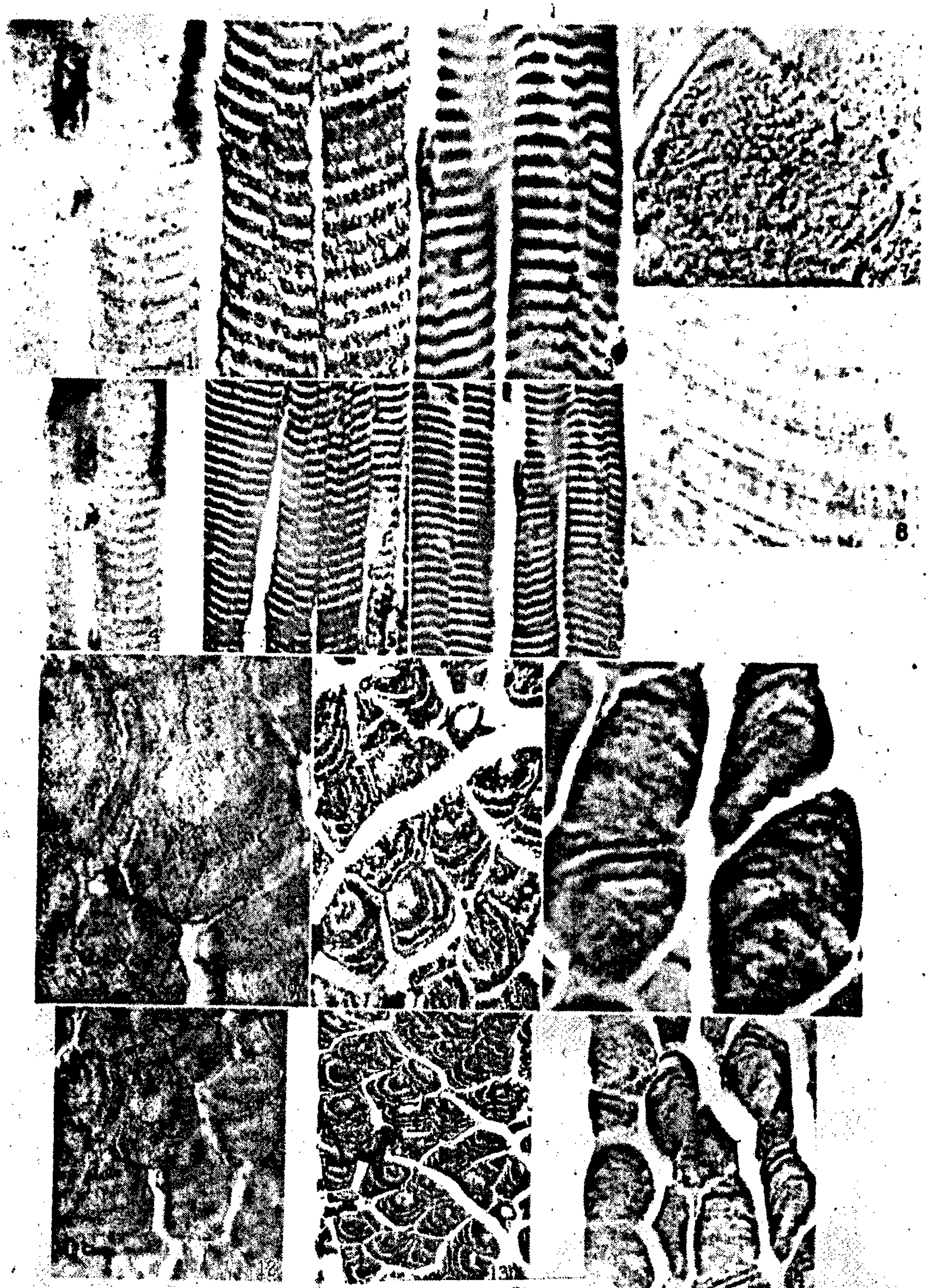


Plate No. 20

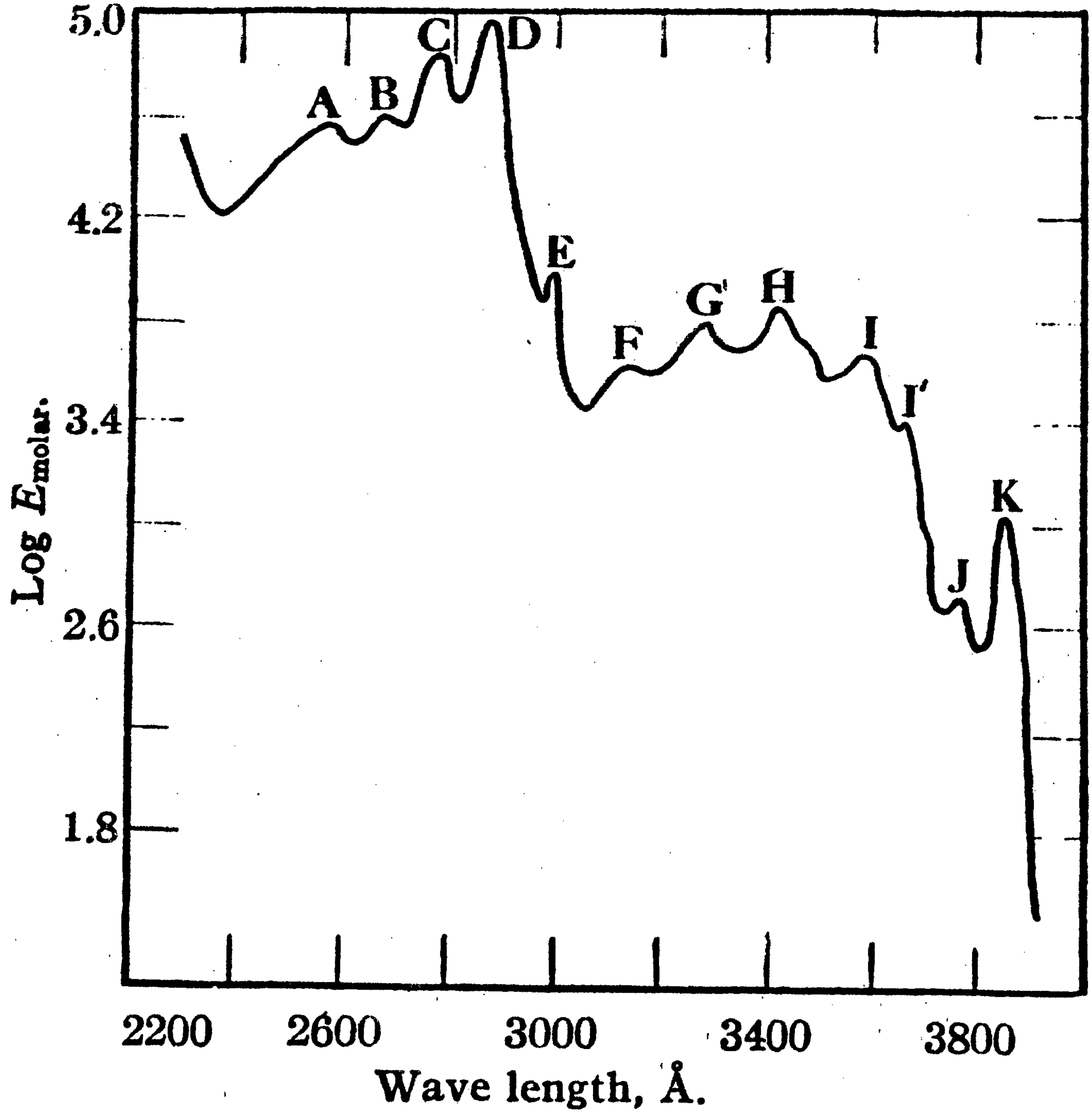


Fig. 1.—1,2-Benzanthracene.

Plate No. 21

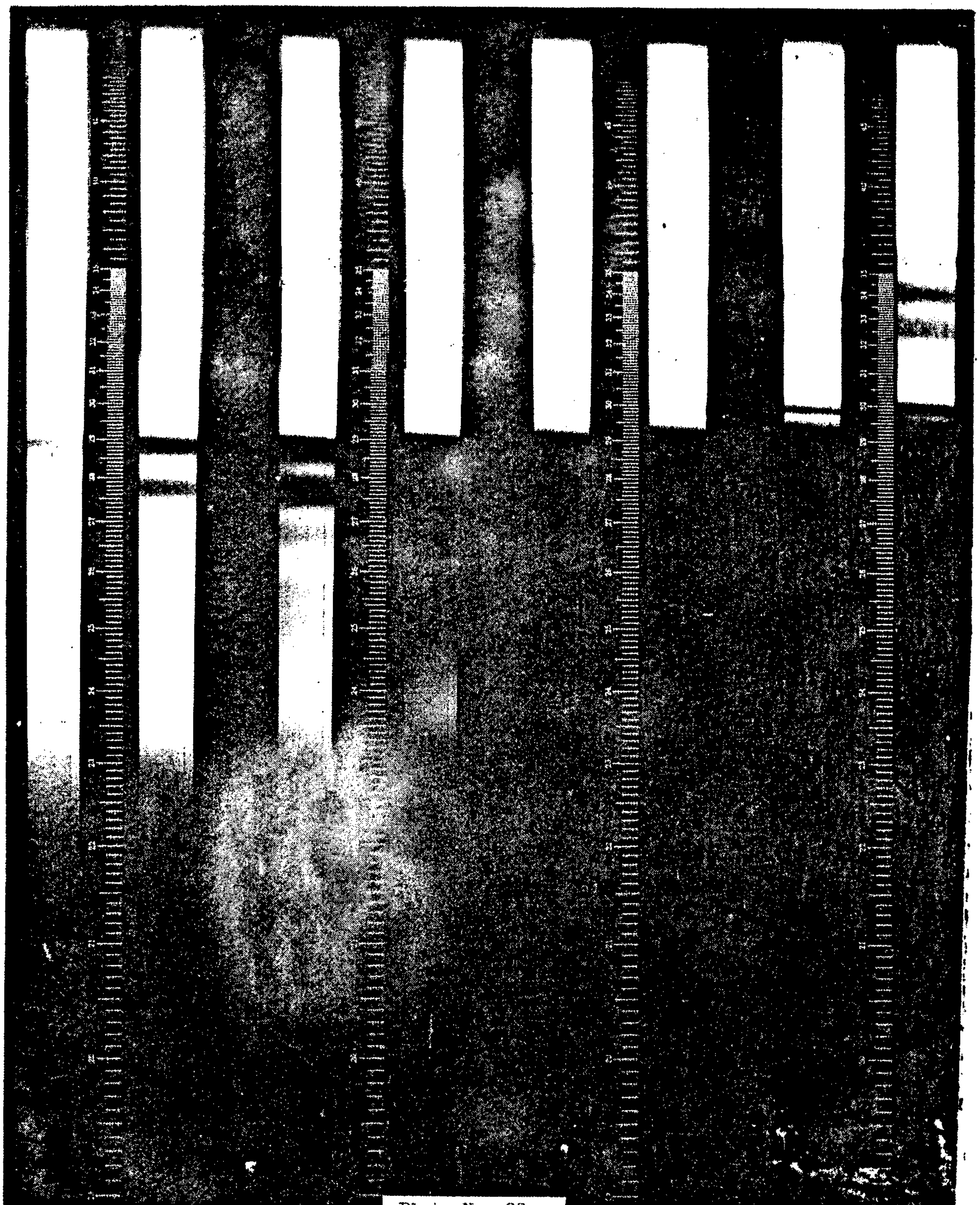


Plate No. 22

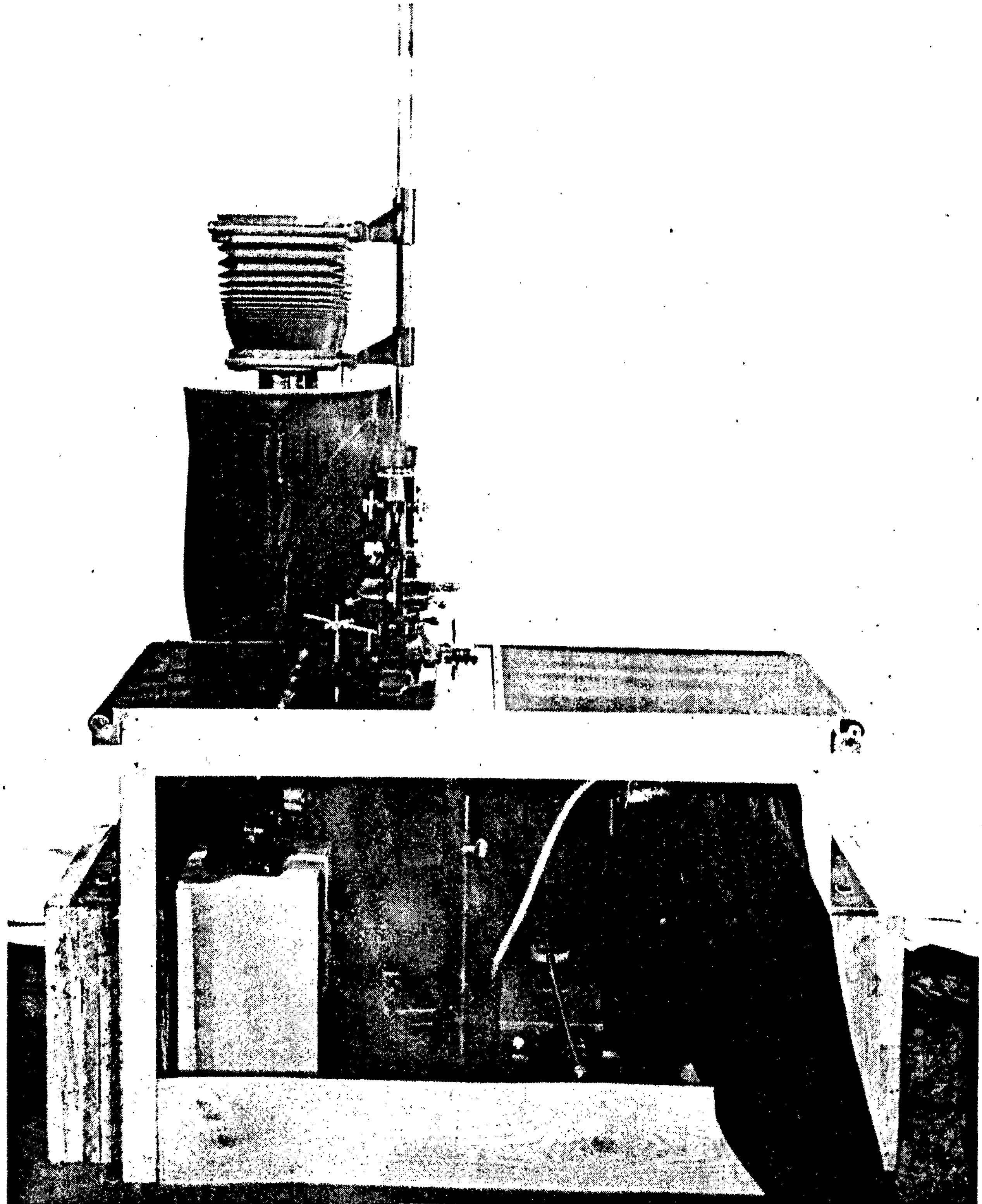


Plate No. 23

REDUNDANCIES IN HUMAN BIOMECHANICS AND THEIR APPLICATION IN ASSESSING MILITARY MAN-TASK DISABILITY PERFORMANCE RESULTING FROM BALLISTIC AGENTS

William H. Kirby, Jr.
Ballistic Research Laboratories

ABSTRACT. Nervous system redundancies have a counterpart in human biomechanics which are evident when analyzing man-task performance using simulations of functional disabilities. These multiple choice pathways available for task accomplishment could explain some of the surprisingly small drops in performance in the presence of otherwise significant tissue trauma especially in cases of highly motivated soldiers. Such findings are thought to be useful in a bioengineering approach for optimizing man-machine or man-task relationships and deriving technical rules for the design of body armor and protective clothing for minimizing personal injury from ballistic agents.

INTRODUCTION. This is a discussion concerning a problem which associates engineering and medicine in a military context. As such it is a cross-application of exact and inexact sciences. With the advent of more intensive interest in mathematical applications to biological problems and the maturation of systems analysis, one now has more formidable tools with which to attack such problems. However, this discussion is restricted to an identification of the natural conditions allowing for multiple choice pathways in task accomplishment. No mathematical model is proposed.

Engineering and mathematical applications to systems behavior, particularly "living-systems", have brought about many new ideas regarding complex problem analysis. An effect of this has been and is, increasingly, that of providing us with more understanding of the nature of these systems. This encourages us to go back to our natural systems for re-examination but with more enlightenment on what we are really looking for.

This morning attention is directed to the response of complex living systems, specifically, humans, under a very large spectrum of traumatic or wounded states. Such interactions are of interest in a setting of military stress situations. Furthermore, it is important to correlate these responses with performance associated with defined tasks (or military occupational specialties) as they appear in the various tactical roles.

In our first approximations, we assume the presence of sufficiently high motivation such that if, despite the presence of trauma, a human biomechanical capability to perform in any fashion exists, it will be used. Consideration today will be restricted to qualitative structure-function relationships and useful methods of analysis. There is significant starting data relating to gross mechanisms for uninjured human function in terms of anatomical and physiological knowledge. One useful approach to this kind of problem is the use of simulation procedures and gaming. This could give us some idea of the value of a loss of function regardless of cause. Obviously wounds cannot be directly simulated, but many endpoint effects in terms of functional disabilities can be. Finally, another essential is the ability to study man-task interactions as they relate to accepted performance levels. In this respect industrial engineers, ergonomists, and applied experimental psychologists are evolving much useful information.

DISCUSSION. Ballistic agents directed at potential enemy soldiers are assessed in terms of the incapacitating effects which they bring about. A wound per se may have little meaning in a military stress situation unless it is matched against task requirements. Incapacitation is defined in this frame of reference as the "diminished capability to perform a defined task." Specifically, a wounded enemy soldier may have an otherwise serious wound in an upper limb, but should his task (military occupational specialty) require the use of only one limb to handle his weapon or machine for counter use during the stress engagement, and thereby effectively incapacitate his enemy (us), then the incapacitating effect of our ballistic agent is not great. It may not even be significant. Of course, the converse may be true. A small but deeply penetrating wound may be very incapacitating. An example of the latter could be one in which a soldier is required to have extremely fine motor coordination in both limbs simultaneously in order to perform the task adequately. In this case, quite a large variety of small penetrating wounds at various anatomical locations could be cited in which motor (muscle) control would be lost resulting in either an inability to control fine movement or an inability to manipulate the entire limb into position for accomplishing the fine movement. The point, of course, in this definition of incapacitation is the fact that an evaluation concerns, not man alone, but man coupled to his assigned tasks in prescribed environments.

As physicians we think in terms of pathological mechanisms pertaining to injury. Such understanding is predicated, in turn, on knowledge

from the basic medical sciences. Functional disabilities ascribable to any spectrum of injuries are assumed to be variable and time dependent. Functional deficits are related in this human system analysis to the final common motor pathway which is a clinical way of referring to one's reactive muscle mechanics. This applied concept is useful inasmuch as the interlock existing between man and his external physical mission or mechanical work is to be accomplished is one of dynamic character. At times this interlocking dynamics is characterized by numerous physical actions while at other times, such actions are numerically minimal but the physiological requirements are intense. In any event, it is obvious that in studying human actions we must deal with a complex integration of both physical and physiological factors.

As engineers we express work in terms of force times distance. Physiologically we have often considered work on the part of a living system as an expenditure of metabolic energy. In this sense we are, of course, in error in our definition of work. We need to differentiate between metabolic energy utilized for life support or maintenance of homeostatis and that metabolic energy available and employed to accomplish "useful" external work. In engineering we sometimes use the expression "useful work" when discussing thermal efficiency of a machine. For example, the thermal efficiency of a locomotive may be 8 per cent, meaning that 8 per cent of the available energy is converted to so many foot-pounds of output and the remainder to heat. However, in handling the combination of inanimate and animate (man and weapon), some practical criteria for NORMAL man-task behavior are needed against which one can assess performance for incapacitated states.

In our first approximations, it appears practical to use a qualitative form of biomechanics providing we can correlate this satisfactorily with appropriate performance levels as derived from experimental analysis. Anthropometrical, physiological, and psychological variations in the human are, of course, an integral part of any appraisal of human function and behavior in the final analysis. Hopefully one looks forward to mathematical expressions for causal relations that are valid and reliable.

Man is endowed with considerable versatility. The use of this versatility is influenced by the mind (brain), body, and environment. The need to relate wound tract pathology with the internal biomechanics of body actions is apparent. However, we also find it useful to study overall functional behavior while in action in order to get impressions of the

sequential dynamics that differentiate one kind of task from another. These applied human mechanics can be studied in terms of their mechanistic factors ... i. e., bone-muscle function. In the long run our model must include input factors such as sensing and/or kinesthetic items as well as kinematics, time-varying factors such as endurance and fatigue, and system alterations due to environment in this total biological-task-environmental stress system problem.

The biophysics and pathological dynamics of wounding are matters of great importance but beyond our scope this morning. We will attempt here to explore the problem in reverse. We may begin with certain functional (motor) deficits which could conceivably be derived from one or more wounds. In the future we expect to assess other factors in the human circuitry such as loss of sensory function, kinesthetic changes, and trauma to the body subsystems such as the nervous system, cardio-vascular system, and respiratory system. We hope to be able to draw from others doing human research in the life sciences and biotechnology. As mentioned, we will try to relate traumatic effects to that final common motor pathway. By simulating motor deficits we are, in effect, working backwards. The more general motor deficits are not difficult to simulate and as such enable us to make at least first approximations of man-task (soldier-weapon or soldier-task) performance behavior as this occurs in simulated tactical role exercises. Incidentially, it is felt that this approach will allow us to do as Professor Pearson outlined yesterday ... ...weld together the important elements including mathematical and war and computer games on the one hand and laboratory experiments, range trials, and Army exercises on the other.

Our view of a human-task model includes the following factors:

1. The Human or Living Machine
2. Tasks
3. Interactions of the Living Machine and Task Accomplishment
4. Environment (not discussed in this paper)

The Human or Living Machine.

In our present work, the human body is being considered as a living machine endowed with a natural structural system capable of serving its internal biological needs while performing external tasks. Many sources

are available for descriptions of human mechanics. The Hall of Biology of Man of the American Museum of Natural History is one of these. Anatomists, particularly J. C. Grant (A Method of Anatomy) and G. B. Duchenne (The Physiology of Motion, more recently translated by E. B. Kaplan of Columbia University), offer very detailed work along these lines. Your attention may have been attracted recently to the present series of articles in LIFE Magazine which discusses some of the biomechanics of the human body.

While many people have not thought about human function in terms of mechanics, they are not particularly surprized when human activity is conservatively compared with mechanical devices. It is also not difficult to see that some overlap or redundancy must exist in a system in which the parts are so intricately intertwined as in the human.

Studies of normal (non-injured) human subjects performing given tasks show that different people use different methods in performing them. In addition, these methods are modified as subjects become more accustomed to the tasks. We say that workers become conditioned or skilled in their tasks. Handicapped workers are often forced to use even radically different methods. In the few simulated incapacitation studies that we have made, we have found a considerable variety of methods employed by our limited number of subjects. In a real sense, then, we can consider this multiplicity of methods in terms of biomechanical redundancies. As a matter of fact, one can often explain the adaptation or describe the compensatory pathway chosen on the part of a disabled individual when a more natural one is not available for performing a given task.

Tasks.

Turning briefly to the applied or task side of the picture, i.e., the nature of the things that impose motor output requirements on the human machine, we find an almost endless array of situations. Knobs have to be grasped, turned, and released; levers (such as rifle trigger, bolt, etc.) have to be pulled and pushed; things have to be grasped, raised, lowered and released; buttons have to be pushed, tools have to be handled, vehicles guided, etc. All such physical tasks may be arbitrarily viewed as being accomplished by composites of elementary actions performed by the human (machine).

Interactions of the Living Machine and Physical Tasks.

Viewing the human machine in action shows a link-linkage system including its affixed but integrated motors which raise and lower limbs (for walking, lifting objects, etc.), raise, lower, turn, and tilt the head, etc. Identification of these muscle-motor systems, their sequence(s), and associated force factors can be studied extensively. Muscle and joint functions are quite well known. Their motion and force vectors can be studied. The kinematic processes activated by the human motors can be at least approximated by various methods and are very informative from both qualitative and quantitative points of view. Data are being generated in many of the associated sciences involved in the type of cross-application presented here, and the ways and means are evolving rapidly for handling such massive bits and pieces of acquired information. The need for the systems analysis has already been mentioned. Information handling, cybernetic modeling using acceptable analogs, computer technology, statistical and adaptive control techniques are included in the knowledge sources useful in understanding complex system behavior.

Today I would like to call attention to a very crude experiment which indicates the multiplicity of motor pathways available to the human in performing given tasks. Variations in weapon design, man-task functional modes, objective definitions, parameters and boundary conditions have not been specified from a systems viewpoint. We have used very simple immobilization techniques for inhibiting certain motor functions in order to enforce and observe the use of alternative motor pathways. This experimental exercise has to do with rifle firing and reloading. The target is about 2 feet by 2 feet at a distance of approximately 40 yards on a flat terrain. One of the first questions of interest to us was, "Can one perform under functional disability conditions, rather than how well"? However, as you will see from the film, we get cues as to how well one does and can perform.

In order to be a little more inclusive, the rifle experiment was run using two different weapons, namely, the U. S. M-14 and the Russian AK both of which are standard items. We will not speculate on weapon differences as such, but we will mention a few things about a man-rifle relationship with and without functional losses. You will observe in the film that a soldier can fire his weapon quite effectively in all firing positions without either upper limb. Pistol grips and monopods seem to be quite helpful to a soldier so disabled. Now let us view the film.

A MAN-TASK DISABILITY EXERCISE: RIFLE FIRING. Time will allow for only a few general remarks. I would like to say that the more one reviews and studies motion pictures of man-task behavior the more one can comprehend the biomechanics taking place in such exercises. The anatomist may be the first to observe the variations in anatomical mechanics; the statistician may quickly detect probabilities in terms of cause and effect relationships; the abstract mathematician may see early cues for a stochastic model; the mechanical engineer may be the first to note the vectorial mechanics in 3-dimensional space; the physiologist may immediately observe the abrupt discontinuities in the human functional activities and feel more sensitive about the corresponding metabolic requirements involved; etc.

At this time, we have no valid statistical data. In our very crude investigation, we have observed with caution, of course, the following:

1. Partial losses up to and including either total limb did not prevent the subjects from firing and maneuvering.

2. Target scores did not decrease much below the 65% - 75% accuracy range even for the most severe simulated disability that we employed . . . the inability to use the trigger arm.

3. Firing rates and reload times were only a few seconds longer in the absence of a total upper limb.

4. The motions exhibited by the subjects are somewhat influenced by size, shape, and weight characteristics of the weapons. However, required body positions. . . . . .standing, prone, etc. . . . naturally influence the biomechanical adjustments which must take place.

MOTION STUDY OF HUMAN ACTION. Since human work is accomplished by means of body actions, the study of body movements has evolved as one of the principal approaches to the problem of finding more effective ways of performing tasks. From such empirical studies, rules have been developed which are available for more effective application in planning and designing tasks, machines, and weapons. Skillful application of these principles diminishes fatigue. It is interesting but, perhaps, not surprising that the methodology proposed for the study of incapacitation i. e., functional deficit simulation, is useful in studying normal man-task phenomena.

In recent years, physicians, applied experimental psychologists, physiologists, and engineers have been studying performance factors and have developed considerable empirical data especially in regard to environmental perception and functional response. Others giving much thought to human anatomical function include (in addition to anatomists) orthopedists, physical therapists, and designers of prostheses. Functional anthropologists, physicists and advanced systems engineers are also becoming essential participants in this area of activity. However, we have not fully recognized the capabilities of these professions in cross-discipline or cross-professional applications. A sizeable effort has been directed in more recent years, to a so-called ergonomic approach to man-task analysis especially in Europe. This approach places more emphasis on the integration of anatomy, psychology and physiology as well as economics for solving problems in human performance.

The Gilbreths, pioneers in the development of human motion principles, devised a list of 17 so-called "elements" or "therbligs" as they have been called. Such elements have been considered as basic units of motion and apply for the most part to the upper limb functions. They include such terms as "transport empty" meaning moving the hand from one position at a work place to another in that vicinity; "transport loaded" meaning the same thing except that the hand is now carrying an object in which case the characteristics of the object are specified; "grasp" meaning a securing of an item either by a pinch-grasp performed by the fingers or a palmar-grasp as performed by enclosing the hand about the item; "hold" meaning that the hand in question is maintaining an object in a fixed position while, perhaps, the opposite hand is doing something to the object as may occur in an assembly operation; etc. In each case, the elements are timed. Therbligs are usually measured to one-thousandth of a minute. Stop-watch time and motion study engineers usually measure their elements in terms of one-hundredth of a minute. Instead of using therbligs, they use descriptive terms which are more general, such as "pick up hammar", "tap dowel to flush position", "place hammar aside", "place assembly in tote box", and "measure outside diameter with micrometer". As long as the the motions are defined and measured consistently, they are useful in the analysis and the synthesis of operations.

Very quickly I would like to show you the motion study and analysis technique for a given operation which is described on the following slides (figures). Our slides show several motion study charts based on a hypothetical exercise. In slide 1 (Figure 1) we show a "Right-hand: Left-hand Chart" and for simplicity only the actions of the hands are described.

MOTION ANALYSIS - STEP I

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND
1. *HAND* ASSISTS IN BRINGING THE RIFLE (REAR STOCK) TO RIGHT SHOULDER WHILE *INDEX FINGER* IS LOOSELY POSITIONED ON THE TRIGGER. THE *HAND* MAINTAINS A MODERATELY FIRM GRIP ON THE STOCK.	1. *HAND* ASSISTS IN BRINGING RIFLE (FRONT STOCK) TO THE SHOULDER LEVEL, WITH *PALM* AND *FINGERS* MAINTAINING A FIRM SUPPORT AND GRASP.
2. THE REAR STOCK IS POSITIONED AGAINST THE FRONT OF THE SHOULDER; THE *HAND* INCREASES ITS GRIP ON THE STOCK EXCEPT FOR THE *INDEX FINGER* WHICH MAINTAINS AN ESSENTIALLY NEUTRAL POSITION.	2. THE FRONT STOCK IS POSITIONED IN CONJUNCTION WITH THE REAR STOCK AND THE *HAND* TAKES ON A MODERATELY SEVERE SUPPORT FUNCTION.
3. WHILE THE *HAND* (AND *FINGERS*, EXCEPT THE *INDEX FINGER*) MAINTAINS A FIRM GRIP ON THE REAR STOCK, THE *INDEX FINGER* PRESSES AGAINST THE TRIGGER, INCREASING THE PRESSURE UNTIL IT TRIPS THE FIRING PIN.	3. *HAND* SUPPORTS THE FRONT STOCK.
4. IMMEDIATELY AFTER ABSORTION OF THE RECOIL, THE *INDEX FINGER* IS SHIFTED (MEETING NO RESISTANCE) TO THE ORIGINAL POSITION OF THE NEXT ROUND.	4. *HAND* SUPPORTS THE FRONT STOCK.
5. SAME AS 3.	5. *HAND* SUPPORTS THE FRONT STOCK.
6. SAME AS 4.	6. *HAND* SUPPORTS THE FRONT STOCK.
7. SAME AS 3.	7. *HAND* SUPPORTS THE FRONT STOCK.
8. SAME AS 4.	8. *HAND* SUPPORTS THE FRONT STOCK.
9. SAME AS 3.	9. *HAND* SUPPORTS THE FRONT STOCK.
10. SAME AS 4.	10. *HAND* SUPPORTS THE FRONT STOCK.
11. SAME AS 3.	11. *HAND* SUPPORTS THE FRONT STOCK.
12. SAME AS 4.	12. *HAND* SUPPORTS THE FRONT STOCK.
13. SAME AS 3.	13. *HAND* SUPPORTS THE FRONT STOCK.
14. SAME AS 4.	14. *HAND* SUPPORTS THE FRONT STOCK.
15. OPPOSITE OF 1.	15. OPPOSITE OF 1.

FIGURE 1

You observe that the wording in this first slide is in lay language and is lengthy, but this was done purposely for our presentation. The motions of each hand are described singly, and then listed synchronously, one beside the other. You will see more clearly in the next slide the relationship of the activities of each hand. However, this first slide does show how motions are described in terms of elements and a method for analyzing them.

In Slide 2 (Figure 2) you see a condensation in the language as compared to Slide 1. This terminology is often employed by motion analysts. Note how quickly you can get a visual time-history of the events on the part of each hand. In order to emphasize the more important factors in such an operation, one usually includes a clear descriptive summary showing numerical relationships between active and inactive motions.

This same operation cycle has been made a little more sophisticated in Slide 3 (Figure 3) by adding a graphical representation for the various kinds of elements. One purpose of this format is to emphasize the idle or inactive elements. I should point out that ordinarily these graphical representations are in terms of time per element giving a quantitative as well as a qualitative measure. Our example here is hypothetical and admittedly would be more effective if a full motion-time study were made. However, the purpose is to point out a methodology useful in studies relating body mechanics to machine and/or task characteristics.

By applying this technique to human tasks and translating such observations into anatomical mechanics, one may be able to specify in more detailed fashion the various biomechanical pathways employed by humans for given functional disabilities. We all know of unusual accomplishments on the part of some handicapped workers. The same is true for some accident victims immediately following trauma. Certainly the time parameter and undoubtedly a host of others are important, such as environmental conditions, task nature, human motivation, etc., and should be specified in modeling man-task behavior under conditions of disability. In this way, redundant bio mechanical networks available to the human may be weighted and probabilitistic methods applied. It is, of course assumed to be necessary to have a definition for normal performance for any given task as a baseline.

MOTION ANALYSIS - STEP I

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND
1. TRANSPORT LOADED	1. TRANSPORT LOADED
2. POSITION	2. POSITION
3. PULL LOADED (INDEX FINGER) & HOLD	3. SUPPORT & HOLD
4. PUSH EMPTY (INDEX FINGER) & HOLD	4. SUPPORT & HOLD
5. PULL LOADED (INDEX FINGER) & HOLD	5. SUPPORT & HOLD
6. PUSH EMPTY (INDEX FINGER) & HOLD	6. SUPPORT & HOLD
7. PULL LOADED (INDEX FINGER) & HOLD	7. SUPPORT & HOLD
8. PUSH EMPTY (INDEX FINGER) & HOLD	8. SUPPORT & HOLD
9. PULL LOADED (INDEX FINGER) & HOLD	9. SUPPORT & HOLD
10. PUSH EMPTY (INDEX FINGER) & HOLD	10. SUPPORT & HOLD
11. PULL LOADED (INDEX FINGER) & HOLD	11. SUPPORT & HOLD
12. PUSH EMPTY (INDEX FINGER) & HOLD	12. SUPPORT & HOLD
13. PULL LOADED (INDEX FINGER) & HOLD	13. SUPPORT & HOLD
14. PUSH EMPTY (INDEX FINGER) & HOLD	14. SUPPORT & HOLD
15. TRANSPORT LOADED	15. TRANSPORT LOADED

SUMMARY

	TOTAL NO. OF ELEMENTS	NO. OF ACTION ELEMENTS	NO. OF PASSIVE SUPPORT ELEMENTS
RIGHT HAND	15	15 (100%)	0 (0%)
LEFT HAND	15	3 (20%)	12 (80%)

FIGURE 2

MAN-TASK CHART

OPERATION: BRING RIFLE TO BEAR ON TARGET FROM THE *PORT ARMS* POSITION TO THE *SHOULDER AIM* POSITION, FIRE SIX ROUNDS, AND RETURN RIFLE TO THE *PORT ARMS* POSITION.

RIGHT HAND	LEFT HAND	RIFLE
TRANSPORT LOADED	TRANSPORT LOADED	IDLE
POSITION	POSITION	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
PULL LOADED & HOLD	SUPPORT & HOLD	INSTANT OF FIRE
PUSH EMPTY & HOLD	SUPPORT & HOLD	IDLE
TRANSPORT LOADED	TRANSPORT LOADED	IDLE

SUMMARY

	TOTAL NO. OF ELEMENTS	NO. OF ACTION ELEMENTS	NO. OF PASSIVE SUPPORT ELEMENTS	NO. OF IDLE ELEMENTS
RIGHT HAND	15	15 (100%)	0 (0%)	0 (0%)
LEFT HAND	15	3 (20%)	13 (80%)	0 (0%)
RIFLE	15	6 (40%)	0 (0%)	9 (60%)

KEY: VERTICAL HATCHING = ACTION ELEMENTS

HORIZONTAL " = SUPPORT "

CROSS " = IDLE "

FIGURE 3

You can see the additional work required to correlate motion mechanics with anatomical causes since we must concurrently consider supporting postural states. The need for unique mathematics is apparent for we are dealing with complex functions. For example, we are interested in the role or contribution of an individual muscle or muscle group in its relationship to an array of essential tasks. This has not been done in the past either by engineers or psychologists to the best of our knowledge for tasks at large.

As stated, it is of primary interest to try and study the relationships that might exist between anatomical structures and their resulting human motion complexes. It becomes obvious that in spite of one or even more than one imposed functional disability, one may perform a defined task satisfactorily. Case studies have been reported where in times of actual injury, humans perform miraculously. Extreme motivation, fear, pain, and a host of other psychological and physiological factors are deeply imbedded. However, if the structure-function pathways did not exist, high motivation alone could not create new physical anatomical entities. Another important spectrum of feasible alternatives derives from the environment not considered in this paper.

In an anatomical review of the upper limb musculo-skeletal system, it is observed that for given muscle dysfunctions, others are often available to assist in the accomplishment of a specified task. Your attention is directed to Table I in which you observe an anatomical matrix of upper limb muscles. The columns are anatomical actions. The rows are the muscles arranged in order from top to bottom to correspond to the proper regions from shoulder to fingers. Notice that in almost all cases, more than one muscle contributes to a given anatomical motion. This is shown by the presence of more than one x in any column.

It is believed that we are getting closer to technical explanations as to why some perform "miraculously" under handicap. It is the motivation that is mirculous, we can do pretty well from the technical or biotechnical point of view.

We are giving serious thought to the development of a functional anatomical-physiological model of man considering all body systems, or subsystems, such as the musculo-skeletal, cardiovascular, nervous, respiratory, gastrointestinal, and genito-urinary, using what physiological feedbacks we know about and can use. Ordinarily we would avoid such

complexity, but you must remember that our analysis actually begins with wound tract pathology or tissue trauma assignable to given fragment physics. A task need not suffer if its imposed demands are not beyond certain limits. Table I shows the structure-function relationships for the human upper limb. Functionally, redundancies are present. This is indicated as mentioned in Table I by examining anatomical functions performed by more than one muscle. In order to be more correct the x's should be of variable size to indicate in a more quantitative sense, weighted contributions on the part of the respective muscles. The point is that redundancies in terms of anatomical capability to perform exists to accomplish external actions. They exist in a given anatomical region. They exist also because of anatomical duality since many tasks are not fully demanding in terms of both limbs simultaneously. Postural changes may compensate satisfactorily for such losses.

RELIABILITY OF SYSTEMS. The reliability of a system's performance may be related to the objective which the system is expected to attain. Since we are concerned with the complex living-system coupled to inanimate components, nature's solutions to reliability phenomena are of special interest. This is especially true if the objective function of a given ballistic fragment is to maximize the disruption of this living-performing arrangement. At the biological-cell level the answer to unreliability is the presence of many more cells than are required, most of which are in parallel linkages. According to W. S. McCulloch, Von Foerster expresses the view that whatever else it (an ordered 'living' system) contributes, a redundancy of structure is fundamental.*,** McCulloch also states . . . "that which is redundant is, to the extent that it is redundant, stable."**

Such phenomena, including information, channels, and structure are important if additional insight is to be attained concerning explanations for the adaptation and compensation powers so often demonstrated by human body behavior.

---

*"On Self-Organizing Systems and Their Environments," by H. Von Foerster. Fron "Self-Organizing Systems Proceedings of an Interdisciplinary Conference." Pergamon Press. 1960.

**"The Reliability of Biological Systems," by W. S. McCulloch, same publication.

Reliability can be defined statistically as the probability "P" that an element or a system will perform satisfactorily for a given period of time. Unreliability is the probability of a failure during a specified time and is given as "q" = 1- P. Here independence assumes that the failure of one element does not affect the probability of failure of any other element. Trauma from various causes has not yet been studied in this sense, but it is being proposed. However, the lack of knowledge of complex biological system behavior would seem to favor the application of stochastic processes at the present time.

If redundancies exist in human biomechanics as well as for the many supporting subsystems, the act of purposely inhibiting any single channel, even if it were feasible to identify it, may not yield applied information most needed at the present time. Other channels, especially those we are not yet aware of, may take over. However, "groups" of redundant channel inhibition may yield significant responses.

If we study task requirements in terms of biomechanical redundancies, we can begin to describe these pathways. It is well known that no task requires all available channels. Hypothetically, then, performance decrement due to a specific channel deficit may not be significant. Our introductory observations regarding given functional losses for defined tasks show some influence on performance behavior. We are probably dealing in some measure with numerical discontinuities and/or non-linear systems.

But what are the boundary conditions? How will these vary from task to task? How much subsystem understanding is required or when can it simply be "blackboxed"? We feel, as others do, that in complex systems in which there is a problem to be solved, it makes sense to speak of the quality of a solution. The method of solution is determined to a large extent by the problem objective. Herein may lie the reason for some of the differences obtained when many views are taken of the same problem. The approach may be identical but inputs may be biased in various ways. Inputs and criteria obviously influence output numbers.

It is interesting to note that research programs in the life sciences are increasingly accentuating the need for increased knowledge concerning fundamental processes, principles, and mechanisms found in biological systems. For the present, we must formulate and use what basic homeostatic mechanisms known by which body processes are regulated. Studies are being made of adaptive and regulatory processes; neural network theory; reflexes and other feedback systems; random redundant processes;

sensing and transducer mechanisms; the specificity of sensory and motor phenomena and other such functions. It has only been in recent years that interest in the study of physical and mathematical principles relating to control systems has become intensified.

Two general approaches to incapacitation assessment are of immediate interest to us, namely, (1) the medical and biophysical aspects of pathological dynamics for certain fragment physics, and, (2) the probable performance alterations for specified tasks due to the traumatic effects of such traumata.

SUMMARY. In summing, it is believed that the following ideas are important:

1. Humans possess a general-purpose type of anatomical structure-function arrangement. Overlap and redundance, indicated by the fact that more than one muscle contributes to the same anatomical function and the presence of duality particularly in the upper and lower limbs, are factors contributing to the nature of this capability. This same idea may be carried down to the cellular level for a single muscle inasmuch as muscle cells exist in a large number of subsets many of which are in parallel. Finally, we might add that even a single living cell may have different capacities depending on many physiological factors including conditioning.

2. In this context it is felt that the load or demand on the general-purpose human is a significant function of task requirements. Quality of performance, imposed physical forces, variable time durations, environmental factors, and motion behavior in 3-dimensional space are items that may vary radically from one task to another.

3. In assessing disability performance a definition of non-disability performance is essential. We recognize the need for unique combinations of talents in this interdisciplinary problem area.

4. A large amount of experimental information concerning man-task dynamics can be generated for specified tasks using conventional work measurement techniques in conjunction with physiological instrumentation and control system knowledge now available.

Appendix

UPPER LIMB MUSCLE CODE

Code	Muscle Name
Shoulder:	
A_1	Trapezius
A_2	Serratus anterior
A_3	Subclavius
A_4	Pectoralis minor
A_5	Pectoralis major
A_6	Subscapularis
A_7	Supraspinatus
A_8	Infraspinatus
A_9	Teres minor
A_{10}	Teres major
A_{11}	Biceps brachii
A_{12}	Coracobrachialis
A_{13}	Triceps brachii
A_{14}	Deltoid
Arm:	
A_{11}	Biceps brachii
A_{13}	Triceps brachii
A_{12}	Coracobrachialis
B_1	Brachioradialis
B_2	Brachialis
B_3	Anconeus
Forearm:	
A_{11}	Biceps brachii
A_{13}	Triceps brachii
B_1	Brachioradialis
B_3	Anconeus
C_1	Supinator
C_2	Pronator quadratus
C_3	Pronator teres
C_4	Flexor carpi radialis
C_5	Extensor carpi radialis longus
C_6	Flexor digitorum sublimis
C_7	Flexor carpi ulnaris
C_8	Extensor carpi radialis brevis
C_9	Extensor carpi ulnaris
C_{10}	Flexor digitorum profundus
C_{11}	Extensor digitorum communis
C_{12}	Palmaris longus
C_{13}	Abductor pollicis longus
C_{14}	Flexor pollicis longus
C_{15}	Extensor indicis proprius
C_{16}	Extensor digiti quinti proprius
C_{17}	Extensor pollicis longus

Code	Muscle Name
Hand:	
C_6	Elexor digitorum sublimis
C_{10}	Flexor digitorum profundus
C_{11}	Extensor digitorum communis
D_1	Extensor pollicis brevis
C_{17}	Extensor pollicis longus
D_2	Abductor pollicis brevis
C_{13}	Abductor pollicis longus
C_{14}	Flexor pollicis longus
D_3	Flexor pollicis brevis
C_{15}	Extensor indicis proprius
C_{16}	Extensor digiti quinti proprius
D_4	Flexor digiti quinti brevis
D_5	Abductor digiti quinti
D_6	Abductor pollicis
D_7	Palmaris brevis
D_8	Opponens pollicis
D_9	Opponens digiti quinti
D_{10}	Lumbricales
D_{11}	Interossei dorsales
D_{12}	Interossei volares

UPPER LIMB MOTOR FUNCTIONS AND CAUSAL FACTORS

	SHOULDER	ARM						FOREARM				HAND AND DIGITS																											
ANATOMICAL REGIONS (WITH CAUSAL FACTORS)		Fl	Ex	Ab	Ad	Ro M	Ro L	Fl	Ex	Su	Pr	Na Fl	Na Ex	Na Ab	Na Ad	Fi Fl 2 pr ph	Fi Fl ds ph	Fi Ex ph	Th Fl ls ph	Th Ex ls ph	Th Fl 2d ph	Th Ab	Fi Ex In	Fi Ad In	Fi Ex 5th	Fi Ex 2 ds ph	Th Ex 2d ph	Th Ab mc	Fi Fl 5th mc	Th Fl mc ph	Th Ad	Fi Fl 5th	Fi Ab 5th	Th Ab mc	Th Ro mc	Fi Ab mc 5th	Fi Fl mc ph	Fi Ro mc 5th	Fi Ab 4	Fi Ad 4
SHOULDER:																																								
A_1	X																																							
A_2	X																																							
A_3	X																																							
A_4	X																																							
A_5	X	X		X																																				
A_6	X	X	X	X	X	X																																		
A_7	X	X		X		X	X																																	
A_8	X			X	X		X																																	
A_9	X				X		X																																	
A_{10}			X		X	X																																		
A_{11}	X		X					X		X																														
A_{12}		X		X																																				
A_{13}			X					X																																
A_{14}		X	X	X																																				
ARM:																																								
A_{11}	X	X						X		X																														
A_{13}			X					X																																
A_{12}		X		X																																				
B_1		X						X																																
B_2								X																																
B_3									X																															
FOREARM:																																								
A_{11}	X	X						X		X																														
A_{13}			X					X																																
B_1		X						X																																
B_2								X																																
B_3									X																															
C_1										X																														
C_2											X																													
C_3											X	X																												
C_4												X		X																										
C_5													X	X	X																									
C_6												X		X	X																									
C_7												X			X																									
C_8													X		X																									
C_9												X			X																									
C_{10}												X				X	X	X																						
C_{11}													X					X																						
C_{12}																						X																		
C_{13}														X	X			X																						
C_{14}																			X		X		X	X																
C_{15}																									X															
C_{16}																											X													
C_{17}														X																										
HAND:																																								
C_8												X			X																									
C_{10}												X				X	X																							
C_{11}													X													X														
D_1														X						X																				
C_{17}														X													X													
D_2																						X																		
C_{13}														X						X		X																		
C_{14}																				X		X							X											
D_3																			X		X							X												
C_{15}																							X	X																
C_{16}																									X															
D_4																																X								
D_5																																	X							
D_6																															X									
D_7																																								
D_8																														X				X	X					
D_9																													X							X		X		
D_{10}																										X											X			
D_{11}																										X											X		X	
D_{12}																										X											X			X

TABLE I

(See Appendix for identification of code)

HALF-NORMAL PLOTS FOR MULTI-LEVEL FACTORIAL EXPERIMENTS

S. A. Krane
C-E-I-R, Inc., Dugway Field Office

1. INTRODUCTION. Half-normal plots for the interpretation of 2^p factorial experiments have been developed and popularized largely through the work of Cuthbert Daniel (see Daniel [1956] and [1959]). In this method the $2^p - 1$ main effects and interactions are estimated from observations on the 2^p treatment combinations. The empirical cumulative distribution of these estimates is then graphically compared with a cumulative distribution derived from a normal population. A rationale for this procedure is found in the approximate normality of the null distribution of the estimates, based upon normality of experimental errors or upon the tendency embodied in the Central Limit Theorem. According to Daniel, the half-normal plot permits the analyst to judge the reality of the largest main effects and interactions and serves to indicate bad values, heteroscedasticity, dependence of variance on mean and some types of defective randomization. The object of the present paper is to indicate and illustrate possible applications of half-normal plots to balance multi-level factorial experiments in general.

2. AN EXAMPLE. It appears easiest to introduce the technique of half-normal plotting for balanced multi-level factorial experiments in the context of a particular example. For this purpose we shall employ Example 8.1 of Davies [1954]. According to the authors (p. 291, "the data . . . are taken from the results of an investigation into the effects on the physical properties of vulcanized rubber of varying a number of factors, the property recorded being the wear resistance of the samples, and the factors being:

A five qualities of filler
B three methods of pretreatment of the rubber
C four qualities of the raw rubber . . ."

The data are reproduced in Table 1. From the data, the author develops the usual analysis of variance as shown in Table 2. The interpretation (Davies [1954, p. 296]) notes the significance of all main effects and two-factor interactions when tested against the three-factor interaction as error.

Table 1. (Table 8.1 of Davies [1954])
DATA OF A 5 x 3 x 4 FACTORIAL EXPERIMENT
WEAR RESISTANCE OF VULCANISED RUBBER

Level of factor A	Level of factor C: 1			2			3			4		
	Level of factor B 1	2	3	Level of factor B 1	2	3	Level of factor B 1	2	3	Level of factor B 1	2	3
1	404	478	530	381	429	528	316	376	390	423	482	550
2	392	418	431	239	251	249	186	207	194	410	416	452
3	348	381	460	327	372	482	290	315	350	383	376	496
4	296	291	333	165	232	242	158	279	220	301	306	330
5	186	198	225	129	157	197	105	163	190	213	200	255

Table 2. (Table 8.16 of Davies [1954])

ANALYSIS OF VARIANCE OF TABLE 8.1

Source of Variation	Sum of squares	Degrees of freedom	Mean square	Variance ratio
Between levels of factor A ..	478,463	4	119,616	374+
B ..	52,794	2	26,397	82.5+
C ..	150,239	3	50,080	156+
Interactions AB	16,807	8	2,101	6.57+
AC	53,890	12	4,491	14.0+
BC	6,416	6	1,069	3.34*
Remainder = interaction ABC ..	7,688	24	320	
Total	766,297	59		

* Denotes significant, that is $\geq$ 5% value but < 1% value.

\+ Denotes highly significant, that is F $\geq$ 1% value.

In order to analyze the given experimental data by half-normal plotting, we shall reduce the data to single degree of freedom sums of squares. The method to be used depends upon the definition of complete sets of orthogonal contrasts for each of the factors A, B, and C. This definition generally is somewhat artibrary, but it is our experience that an experimenter familiar with the nature of the factor levels and the purpose of the experiment can, in most instances, provide sufficient justification for the prior definition of a meaningful complete set of single degree of freedom orthogonal contrasts among the levels. The use of orthogonal polynomials for quantitative levels is often indicated, while for qualitative levels, meaningful comparisons among certain levels are often obvious. On occasion, only a partial set of orthogonal comparisons will appear to be of intrinsic value and it may be necessary to complete the orthogonal set by adding contrasts of no apparent importance. In the absence of useful information on the nature of the levels (except that they are all qualitative) in the present example, we shall be totally arbitrary in defining the contrasts, but will attempt to indicate their potential interpretations. These contrasts are shown in Table 3. For factor A, contrast A_0, is the "null" or "average" contrast*, while A_1 compares the average of levels 1 and 2 against the average of levels 3, 4 and 5, A_2 compares level 1 vs. level 2, A_3 compares level 3 against the average of levels 4 and 5 and A_4 compares level 4 with level 5. The orthogonality of the set is evident in that the coefficients sum to zero for all contrasts except the null contrast and the sum of products of coefficients is zero for all pairs of contrasts. For factor B, the non-null contrasts compare level 1 with level 2 with the average of levels 1 and 3. (In another context, B_1 and B_2 are the orthogonal polynomials for three equally spaced levels, B_1 being the linear contrast and B_2 the quadratic.) For factor C, the contrasts C_1, C_2 and C_3 make the following comparisons among levels, respectively: (1 and 2) vs. (3 and 4), (1 and 3) vs. (2 and 4) and (1 and 4)

---

* Daniel 1962 has suggested the term "null" is inappropriate because of the generally positive expectation of this contrast. We chose the term because (i) it is connoted by our zero subscript notation, (ii) this contrast is not a comparison among levels, and (iii) this contrast is generally "of no consequence" in the analysis.

Table 3. ORTHOGONAL CONTRASTS EMPLOYED

Factor A

Contrast	Level 1	2	3	4	5	Sum of Squares
A_0	+1	+1	+1	+1	+1	5
A_1	+3	+3	-2	-2	-2	30
A_2	+1	-1	0	0	0	2
A_3	0	0	+2	-1	-1	6
A_4	0	0	0	+1	-1	2

Factor B

Contrast	Level 1	2	3	Sum of Squares
B_0	+1	+1	+1	3
B_1	-1	0	+1	2
B_2	-1	+2	-1	6

Factor C

Contrast	Level 1	2	3	4	Sum of Squares
C_0	+1	+1	+1	+1	4
C_1	-1	-1	+1	+1	4
C_2	-1	+1	-1	+1	4
C_3	+1	-1	-1	+1	4

vs. (2 and 3). These contrasts would be of interest, e. g., in the event that factor C incorporated two subfactors, say D and E, where levels 1 and 2 are at the low level of D and levels 3 and 4 at the high level of D, while levels 1 and 3 are at the low level of E and 2 and 4 at the high level of E. Then C_1 is the effect of D, C_2 is the effect of E and C_3 is the interaction of D and E.

The three sets of contrasts (A_1, A_2, A_3, A_4), (B_1, B_2) and (C_1, C_2, C_3) will provide a basis for reducing the sums of squares for factor A (4 d. f.), factor B (2 d. f.) and factor C (3 d. f.) to independent single degree of freedom sums of squares. It remains to develop such a basis for the two- and three-factor interactions. A natural method for accomplishing this is the extension of the original single factor contrast sets to interaction contrast sets. This method is exemplified in Table 4 for Factors B and C. All possible combinations of the levels of B and C are employed as columns, while rows are contrasts. For any combination of a particular level, say i, of B with a particular level, say j, of C, the coefficient in the contrast B_qC_r is obtained by multiplication of the coefficient of level i of B in the contrast B_q by the coefficient of level j or C in the contrast C_r. Sums of squares of the B and C contrasts may be obtained by multiplication of the corresponding sums of squares for B and for C.

Of the 12 orthogonal contrasts in Table 4, B_0C_0 is the null contrast while the contrasts B_0C_1, B_0C_2, B_0C_3, B_1C_0 and B_2C_0 are simply the original contrasts C_1, C_2, C_3, B_1 and B_2, respectively, averaged over all levels of the other factor. The six contrasts B_1C_1, B_1C_2, B_1C_3, B_2C_1, B_2C_2, B_2C_3 are new and constitute a basis for partitioning the BC interaction sum of squares (6 d. f.) into orthogonal single degree of freedom sums of squares. Application of this method will likewise produce bases for partitioning the sums of squares for AB (8 d. f.), AC (12 d. f.) and ABC (24 d. f.).

The above method of defining interaction contrasts is incorporated in the method we employ for calculating half-normal variates by desk

Table 4. ORTHOGONAL CONTRASTS FOR FACTORS B AND C

	Level of C	1			2			3			4			Sum of
	Level of B	1	2	3	1	2	3	1	2	3	1	2	3	Squares
Contrast	B_0C_0	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	12
	B_0C_1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	12
	B_0C_2	-1	-1	-1	+1	+1	+1	-1	-1	-1	+1	+1	+1	12
	B_0C_3	+1	+1	+1	-1	-1	-1	-1	-1	-1	+1	+1	+1	12
	B_1C_0	-1	0	+1	-1	0	+1	-1	0	+1	-1	0	+1	8
	B_1C_1	+1	0	-1	+1	0	-1	-1	0	+1	-1	0	+1	8
	B_1C_2	+1	0	-1	-1	0	+1	+1	0	-1	-1	0	+1	8
	B_1C_3	-1	0	+1	+1	0	-1	+1	0	-1	-1	0	+1	8
	B_2C_0	-1	+2	-1	-1	+2	-1	-1	+2	-1	-1	+2	-1	24
	B_2C_1	+1	-2	+1	+1	-2	+1	-1	+2	-1	-1	+2	-1	24
	B_2C_2	+1	-2	+1	-1	+2	-1	+1	-2	+1	-1	+2	-1	24
	B_2C_3	-1	+2	-1	+1	-2	+1	+1	-2	+1	-1	+2	-1	24

calculator. A sample work sheet for this method is shown in Table 5a and 5b*. Section I of this table is merely a recopying of the original data from Table 1. In this section a column of five numbers represents the observations for the five levels of A over a particular one of the twelve combinations of levels of B and C represented by columns. Section II is computed by operating on these columns with the contrasts A_0, A_1, A_2, A_3, A_4. For example, products of the coefficients of A_0 with the corresponding elements of a particular column are formed and these five products are summed and entered in Section II in the first row of that column. Similarly, sums of products of coefficients of A_1 with corresponding elements of columns are entered in the second row of Section II, and so forth. Thus each element of Section II is formed as a sum of products of coefficients of, say, A_p with corresponding observations and is entered on the (p+1)-th row of the appropriate column.

Section II may be visualized as an aggregation of 20 rows of three elements each, where each row corresponds to a particular contrast of A and a particular level of C. The three elements of each row correspond to the three levels of B. Now Section III is formed from Section II by summing products of coefficients of the B contrasts and corresponding elements of each row of three. Each such sum of products is entered in the corresponding row, with the sum of products from coefficients of B_q entered as the (q + 1)-th element of that row.

Section III may be visualized as comprising four sub-sections of 15 elements each, with the elements of a sub-section corresponding to a particular contrast of AB (i. e., a particular combination of a contrast of A and a contrast of B) and the four sub-sections corresponding to the four levels of C. Section IV is formed from Section III by summing products of coefficients of the C contrasts and identically placed elements from the four corresponding sub-sections. The sum of products is entered in the corresponding place of one of the sub-sections of Section IV, with the sum of products from coefficients of C_r entered in the (r+1)-th sub-section.

---

*This calculation method is essentially the same as that given in Appendix 8G of Davies [1954, pp. 363-6]. In some instances the format in Davies (with the addition of final column of half-normal variates) may be preferred.

Table 5a. COMPUTATION OF SINGLE DEGREES OF FREEDOM

	404	478	530	381	429	528	316	376	390	423	482	550
	392	418	431	239	251	249	186	207	194	410	416	452
I	348	381	460	327	372	482	290	315	350	383	376	496
	296	291	333	165	232	242	158	279	220	301	306	330
	186	198	225	129	157	197	105	163	190	213	200	255
	1626	1766	1979	1241	1441	1698	1055	1340	1344	1730	1780	2033
	728	948	847	618	518	489	400	235	232	705	930	344
II	12	60	99	142	178	279	130	169	196	13	66	98
	214	273	362	360	355	525	317	188	290	252	246	407
	110	93	108	36	75	45	53	116	30	88	106	75
	5371	353	-73	4380	457	-57	3739	289	281	5593	353	-253
	2523	119	321	1625	-129	-71	867	-168	-162	2479	139	-311
III	171	87	9	599	137	-65	495	66	12	177	85	21
	849	148	-30	1240	165	-175	795	-27	-231	905	155	-167
	311	-2	-32	156	9	69	199	-23	149	269	-13	49
	19083	1452	-102	-419	-168	158	863	168	-518	2845	-40	-550
	7494	-39	399	-802	-19	-101	714	59	81	2510	555	865
IV	1442	375	-23	-98	-73	89	110	69	-65	-746	-31	83
	3789	441	-603	-389	-185	-193	501	199	-81	-281	165	209
	935	-29	235	1	-43	161	-85	21	1	225	-1	-201

Table 5b. COMPUTATION ON SINGLE DEGREES OF FREEDOM (Cont'd)

	$A_0B_0C_0$	$A_0B_1C_0$	$A_0B_2C_0$	$A_0B_0C_1$	$A_0B_1C_1$	$A_0B_2C_1$	$A_0B_0C_2$	$A_0B_1C_2$	$A_0B_2C_2$	$A_0B_0C_3$	$A_0B_1C_3$	$A_0B_2C_3$
	$A_1B_0C_0$	$A_1B_1C_0$	$A_1B_2C_0$	$A_1B_0C_1$	$A_1B_1C_1$	$A_1B_2C_1$	$A_1B_0C_2$	$A_1B_1C_2$	$A_1B_2C_2$	$A_1B_0C_3$	$A_1B_1C_3$	$A_1B_2C_3$
V	$A_2B_0C_0$	$A_2B_1C_0$	$A_2B_2C_0$	$A_2B_0C_1$	$A_2B_1C_1$	$A_2B_2C_1$	$A_2B_0C_2$	$A_2B_1C_2$	$A_2B_2C_2$	$A_2B_0C_3$	$A_2B_1C_3$	$A_2B_2C_3$
	$A_3B_0C_0$	$A_3B_1C_0$	$A_3B_2C_0$	$A_3B_0C_1$	$A_3B_1C_1$	$A_3B_2C_1$	$A_3B_0C_2$	$A_3B_1C_2$	$A_3B_2C_2$	$A_3B_0C_3$	$A_3B_1C_3$	$A_3B_2C_3$
	$A_4B_0C_0$	$A_4B_1C_0$	$A_4B_2C_0$	$A_4B_0C_1$	$A_4B_1C_1$	$A_4B_2C_1$	$A_4B_0C_2$	$A_4B_1C_2$	$A_4B_2C_2$	$A_4B_0C_3$	$A_4B_1C_3$	$A_4B_2C_3$
	60	40	120	60	40	120	60	40	120	60	40	120
	360	240	720	360	240	720	360	240	720	360	240	720
VI	24	16	48	24	16	48	24	16	48	24	16	48
	72	48	144	72	48	144	72	48	144	72	48	144
	24	16	48	24	16	48	24	16	48	24	16	48
	6069348	52708	87	2926	706	208	12413	706	2236	134900	40	2521
	156000	6	221	1787	2	14	1416	15	9	17500	1283	1039
VII	86640	8789	11	400	333	165	504	298	88	23188	60	144
	199396	4052	2525	2102	713	259	3486	825	46	1097	567	303
	36426	53	1151	0	116	540	301	28	1	2109	0	842
	2464	230	-9	-54	-27	14	111	27	-47	367	-6	-50
	395	-2	15	-42	-1	-4	38	4	3	132	36	32
VIII	294	94	-3	-20	-18	13	22	17	-9	-152	-8	12
	447	64	-50	-46	-27	-16	59	29	-7	-33	24	17
	191	-7	34	0	-11	23	-17	5	1	46	-0	-29

The method of calculation of Sections II, III and IV may not be apparent at first glance, but verifying part or all of the data in Table 5a from the description above should help to clarify the process. Computing clerks will find it helpful to write the coefficients of each contrast on a strip of paper, appropriately oriented vertically or horizontally and spaced so that when overlaid on the worksheet each coefficient appears adjacent to the element to be multiplied.

Section V (Table 5b) merely identifies the elements of Section IV and subsequent sections according to the contrasts they represent. This identification is, of course, highly systematic and might well be omitted when familiarity with the method is attained.

Section VI contains the "divisors", obtained by multiplying the sums of squares of the coefficients of the contrasts A_p, B_q, C_r appropriate to each element, as found in Table 3.

Section VII contains the single degree of freedom sums of squares corresponding to each contrast. Each element is obtained by squaring an element of Section IV, dividing by the corresponding element of Section VI and entering in the corresponding place of Section VII.

Section VIII contains the half-normal variate values, each of which is computed as the square root of the corresponding element of Section VII, positive or negative according to the sign of the corresponding element of Section IV. (It would perhaps have been advisable to include the first decimal of each of these values in order to discriminate more fully among them.)

Certain check computations in the method have been omitted, but an over-all check can be readily obtained from Section VII by comparing sums of these single degree of freedom sums of squares with the usual analysis of variance of Table 2. These checks are indicated in Table 6. It will be noted that all sums of squares agree with Table 2 within the expected rounding error accumulated from Section VII.

The half-normal variates must now be ordered by magnitude before plotting. This ordering is shown in Table 7, along with an identification of the contrast represented (letters with subscripted zeroes have been dropped) and the appropriate quantile of the empirical distribution, defined by

Table 6. DEVELOPMENT OF USUAL ANALYSIS OF VARIANCE

Source	d.f.	Contrasts	S.S.
A	4	$A_1B_0C_0$ $A_2B_0C_0$ $A_3B_0C_0$ $A_4B_0C_0$	478462
B	2	$A_0B_1C_0$ $A_0B_2C_0$	52795
C	3	$A_0B_0C_1$ $A_0B_0C_2$ $A_0B_0C_3$	150239
AB	8	$A_1B_1C_0$ $A_2B_1C_0$ $A_3B_1C_0$ $A_4B_1C_0$	
		$A_1B_2C_0$ $A_2B_2C_0$ $A_3B_2C_0$ $A_4B_2C_0$	16808
AC	12	$A_1B_0C_1$ $A_2B_0C_1$ $A_3B_0C_1$ $A_4B_0C_1$	
		$A_1B_0C_2$ $A_2B_0C_2$ $A_3B_0C_2$ $A_4B_0C_2$	
		$A_1B_0C_3$ $A_2B_0C_3$ $A_3B_0C_3$ $A_4B_0C_3$	53890
BC	6	$A_0B_1C_1$ $A_0B_1C_2$ $A_0B_1C_3$	
		$A_0B_2C_1$ $A_0B_2C_2$ $A_0B_2C_3$	6417
ABC	24	$A_1B_1C_1$ $A_2B_1C_1$ $A_3B_1C_1$ $A_4B_1C_1$	
		$A_1B_1C_2$ $A_2B_1C_2$ $A_3B_1C_2$ $A_4B_1C_2$	
		$A_1B_1C_3$ $A_2B_1C_3$ $A_3B_1C_3$ $A_4B_1C_3$	
		$A_1B_2C_1$ $A_2B_2C_1$ $A_3B_2C_1$ $A_4B_2C_1$	
		$A_1B_2C_2$ $A_2B_2C_2$ $A_3B_2C_2$ $A_4B_2C_2$	
		$A_1B_2C_3$ $A_2B_2C_3$ $A_3B_2C_3$ $A_4B_2C_3$	7690
Total	59		766301
Mean	1	$A_0B_0C_0$	6069348
Raw total	60		6835649

Table 7. HALF-NORMAL VARIATES

Order k	Variate X_k	Contrast	Quantile P_k	Order k	Variate X_k	Contrast	Quantile P_k
60	2464	Null		30	24	$A_3B_1C_3$	.5000
59	447	A_3	.9915	29	23	$A_4B_1C_1$	.4831
58	395	A_1	.9746	28	22	A_2C_2	.4661
57	367	C_3	.9576	27	20	$-A_2C_1$	.4492
56	294	A_2	.9407	26	18	$-A_2B_1C_1$	.4322
55	230	B_1	.9237	25	17	$A_3B_2C_3$	.4153
54	191	A_4	.9068	24	17	$-A_4C_2$	.3983
53	152	$-A_2C_3$	.8898	23	17	$A_2B_1C_2$	.3814
52	132	A_1C_3	.8729	22	16	$-A_3B_2C_1$	.3644
51	111	C_2	.8559	21	15	A_1B_2	.3475
50	94	A_2B_1	.8390	20	14	B_2C_1	.3305
49	64	A_3B_1	.8220	19	13	$A_2B_2C_1$	.3136
48	59	A_3C_2	.8051	18	12	$A_2B_2C_3$	.2966
47	54	$-C_1$	.7881	17	11	$-A_4B_1C_1$	.2797
46	50	$-A_3B_2$	.7712	16	9	$-A_2B_2C_2$	.2627
45	50	$-B_2C_3$	.7542	15	9	$-B_2$	.2458
44	47	$-B_2C_2$	.7373	14	8	$-A_2B_1C_3$	.2288
43	46	A_4C_3	.7203	13	7	$-A_4B_1$	.2119
42	46	$-A_3C_1$	.7034	12	7	$-A_3B_2C_2$	.1949
41	42	$-A_1C_1$	.6864	11	6	$-B_1C_3$	.1780
40	38	A_1C_2	.6695	10	5	$A_4B_1C_2$	.1610
39	36	$A_1B_1C_3$	.6525	9	4	$A_1B_1C_2$	.1441
39	34	A_4B_2	.6356	8	4	$-A_1B_2C_1$	.1271
37	33	$-A_3C_3$	.6186	7	3	$-A_2B_2$	.1102
36	32	$A_1B_2C_3$	.6017	6	3	$A_1B_2C_2$	.0932
35	29	$-A_4B_2C_3$	.5847	5	2	$-A_1B_1$	.0763
34	29	$A_3B_1C_2$	.5678	4	1	$-A_1B_1C_1$	.0593
33	27	$-A_3B_1C_1$	.5508	3	1	$A_4B_2C_2$	.0424
32	27	$-B_1C_1$	.5339	2	0	$-A_4B_1C_3$	.0254
31	27	B_1C_2	.5169	1	0	A_4C_1	.0085

$$P_k = \frac{2k - 1}{2n}$$

where k is the rank order and n is the number of variates. Here, as in most instances, it seems appropriate that the null contrast be excluded from the variates to be examined. The sign of the contrast is now attached to the label and only positive variates are plotted.

The variate values and quantiles are next plotted on half-normal probability paper (as in Figure 1) for interpretation. Discussion of the interpretation phase of the analysis of this example will be deferred to a later section.

3. SOME THEORY*. At this point we shall touch briefly on some theoretical aspects of the development of half-normal variates from multi-level factorial experiments. To simplify the discussion we shall assume that we are concerned with a three-factor experiment, although it should be remembered that the theory and methodology apply with equal validity to any number of factors.

We denote by y_{hij} the observation obtained with factor A at level h, factor B at level i and factor C at level j, where $h = 1, 2, \ldots, a$; $i = 1, 2, \ldots, b$; $j = 1, 2, \ldots, c$. The coefficients of the orthogonal contrasts for factor A will be indicated by a_{ph}, denoting the coefficient for level h in the p-th contrast. Similarly the coefficients of the contrasts for factors B and C are denoted b_{qi} and c_{rj}, respectively. We assume that for each factor there is a null contrast, these being denoted A_0, B_0, C_0 and defined by

$$a_{oh} = b_{oi} = c_{oj} = 1; \text{ all } h, i, j.$$

*This section is based on well-known results concerning distributions of linear functions of random variables and may be verified by reference to standard introductory texts on mathematical and theoretical statistics.

Furthermore, by the definition of orthogonal contrasts,

$$\sum_h a_{ph} = \sum_i b_{qi} = \sum_j c_{rj} = 0$$

$$p = 1, 2, \ldots, a-1; \quad q = 1, 2, \ldots, b-1; \quad r = 1, 2, \ldots, r-1;$$

and

$$\sum_h a_{ph} \; a_{p'h} = \sum_i b_{qi} \; b_{q'i} = \sum_j c_{rj} \; c_{r'j} = 0$$

$$p \neq p'; \quad q \neq q'; \quad r \neq r'.$$

The three-factor contrasts are defined by

$$(A_p B_q C_r) = \sum_h \sum_i \sum_j a_{ph} \, b_{qi} \, c_{rj} \, y_{hij};$$

$$p = 0, 1, \ldots, a-1; \; q = 0, 1, \ldots, b-1; \; r = 0, 1, \ldots, c-1.$$

Suppose that there are no treatment effects*, i.e.,

$$E\{y_{hij}\} = \mu; \text{ all } h, \; i, \; j;$$

and that the experimental errors are independent and have constant variance for all observations, i.e.,

$$E\{(y_{hij} - \mu)^2\} = \sigma^2; \text{ all } h, \; i, \; j.$$

---

*The symbol $E\{\}$ denotes the mathematical expectation operator.

Then

$$E\{A_p B_q C_r\} = 0,$$

unless

$$P = 0, \; q = 0 \text{ and } r = 0, \text{ in which case}$$

$$E\{A_0 B_0 C_0\} = abc\,\mu.$$

Furthermore*,

$$V\{A_p B_q C_r\} = (\sum_h a_{ph}^2)(\sum_i b_{qi}^2)(\sum_j c_{rj}^2)\,\sigma^2.$$

Denote by Y_{pqr} the variate defined by

$$Y_{pqr} = (A_p B_q C_r) \Big/ \sqrt{(\sum_h a_{ph}^2)\,(\sum_i b_{qi}^2)\,(\sum_j c_{rj}^2)}.$$

Then

$$E\{Y_{000}\} = \sqrt{abc}\,\mu;$$

$$E\{Y_{pqr}\} = 0, \text{ unless } p = 0, \; q = 0, \; r = 0;$$

$$V\{Y_{pqr}^2\} = \sigma^2.$$

If the experimental errors are normally distributed, then the Y_{pqr} are normally distributed. (Under fairly weak assumptions the Y_{pqr} will tend to be normally distributed in large experiments even for non-normal distributions of experimental error.) Then the non-negative half-normal variates,

---

*The symbol $V\{\,\}$ denotes the variance operator, $V\{X\} = E\{(X-E\{X\})^2\}$.

$$X_{pqr} = \left| Y_{pqr} \right| ,$$

$$= \sqrt{(A_p B_q C_r)^2 / (\sum_h a_{ph}^2)(\sum_i b_{qi}^2)(\sum_j c_{rj}^2)}; \; p, q, r \neq 0, 0, 0;$$

are indeed distributed according to the half-normal density

$$f(x) = \sqrt{2/\pi \sigma^2} \quad \exp(-X^2/2\sigma^2), \qquad X \geq 0$$

$$X \leq 0.$$

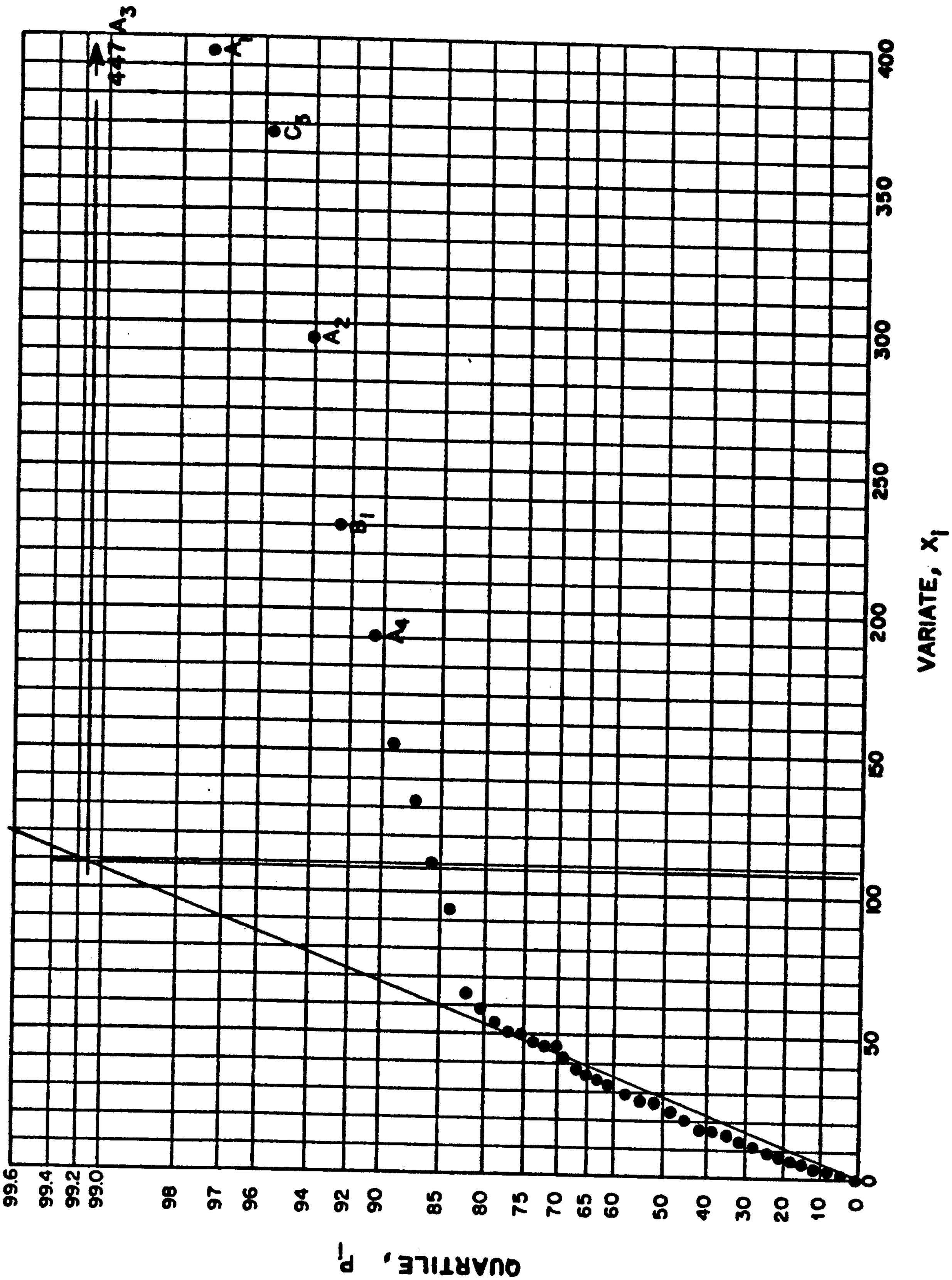
From this result, the half-normal variates for multi-level experiments may be seen to be essentially equivalent to those for 2^p experiments, making the work of Daniel 1959 and Birnbaum 1959 relevant to the interpretation.

4. INTERPRETATION OF EXAMPLE. We shall turn now to the interpretation of the example given earlier. Some difficulty will be experienced because of our ignorance of the precise nature of the factors and their levels, the experimental techniques and the observations themselves, but we shall attempt to proceed along lines suggested by Daniel for 2^p experiments.

To recapitulate the results of Section 2, we have, in Table 7, 59 ordered variates $X_k \equiv X_{pqr}$ whose empirical cumulative distribution should resemble the cumulative half-normal distribution under the hypothesis that there are no treatment effects. We have plotted these values against their quantiles in Figure 1, where they should be approximately linear under the null hypothesis.

We note at a glance that the plotted points are markedly and systematically non-linear. In fact, a little preliminary geometrical construction leads us to believe that a number of the variates are too large to have arisen by chance under the null hypothesis. The rationale for this belief is as follows. Under the null hypothesis the standard deviation, σ, is

FIGURE I. HALF-NORMAL PLOT, ALL CONTRASTS



directly approximated by the value of X_m, where

$$m = (0.683\, n + 0.5),$$

$$= 41, \text{ approximately.}$$

From Table 7,

$$X_{41} = 42.$$

Then, under the null hypothesis, the plotted points should lie near a straight line through the origin and the point (X_m, P_m), indicated in Figure 1. Should the largest X lie "far enough" to the right of this line it is reasonable to presume that it did not arise by chance under the null hypothesis. It may then be taken as real and the next largest X promoted to the largest. This is roughly equivalent to increasing the ordinate of the second point to that of the first point. Should this replotted point also lie "far enough" to the right of the line, it too may be judged real and excluded, promoting the next X to the largest, etc. In Figure 1, we make a crude test of the largest values by constructing a horizontal through the largest point to intersect the previously constructed empirical cumulative distribution line. From this intersection we drop a vertical line and observe that all contrasts represented by points lying to the right of this vertical would have to be excluded before the largest X would lie on or above the original c. d. line. In this crude manner we judge from Figure 1 that six to ten of the largest values of X would be unlikely to occur under the null hypothesis. This graphical construction is no "exact" test; in fact it is rather likely that one or more contrasts would be judged "real" in this manner even if the null hypothesis did, in fact, hold. There is one element of conservativism in this procedure, in that the plotted c. d. line is based upon <u>all</u> contrasts, while a c. d. line based only on contrasts not judged "real" at this stage would lie to the left of the original line.

Let us tentatively suppose that the six largest contrasts (A_3, A_1, C_3, A_2, B_1, A_4) are real, considering (after Daniel [1959, p. 315]) their simple <u>names</u>, as well as their magnitudes relative to the rest of the set. We plot anew the 53 remaining contrasts in Figure 2. Actually, in addition to the ten largest remaining contrasts, only a fraction of the points are plotted, together with the c. d. line through (X_m, P_m), where

$$m = (0.683)(53) + 0.5,$$

$$= 37, \text{ approximately.}$$

The values of P_k are, of course, recalculated for $n = 53$. It appears reasonable to judge from this plot that the four largest contrasts (A_2C_3, A_1C_3, C_2, A_2B_1) are real.

A final plot of the values obtained after eliminating the ten largest values is shown in Figure 3. It appears in this plot that all real effects have been removed, with a residual error standard deviation approximately equal to

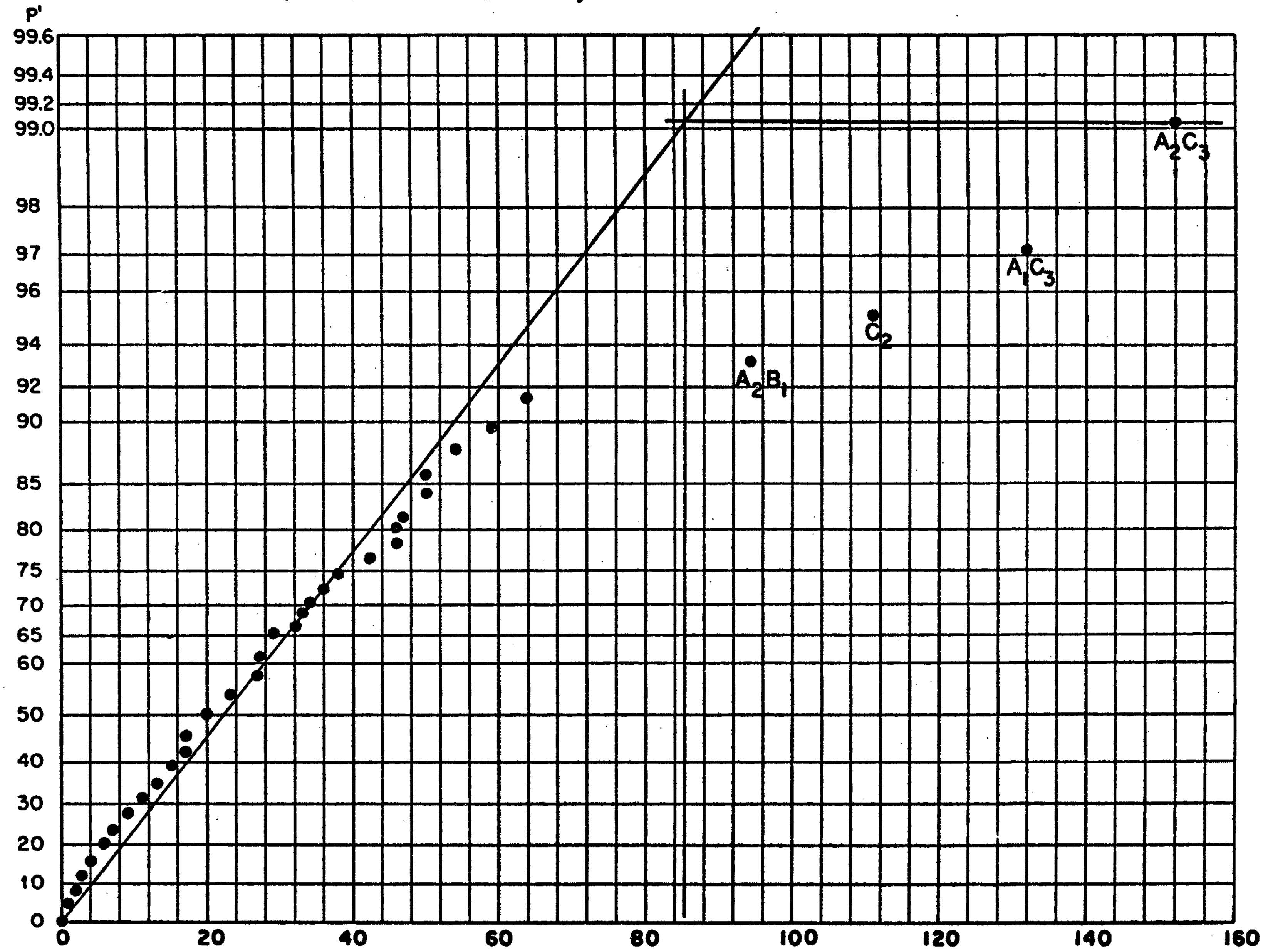
$$X^2_{34} = 841$$

(The actual mean square of the 49 residual contrasts is 816.)

Some further details of interpretation might be attempted. For example, there is a suggestion in Figure 1 and in Table 7 that there may have been plot-splitting, with factor B applied within plots. This also appears plausible from the rudimentary information given as to the nature of this factor. A further plotting, not shown here, in which contrasts including B_1 or B_2 were separated from those containing B_0 suggests a whole plot standard deviation of about 50-60 and a split-plot standard deviation of about 20-25.

5. COMPUTER USE. We have used half-normal plots for multi-level factorial experiments for almost two years. Our first major attempt to employ this technique was in the analysis of an unreplicated 10 x 5 x 3 x 2^2 experiment. The factor levels in this experiment were applied in a split-split-split plot design and certain problems of variance heterogeneity were apparent. The half-normal plotting of this data was sufficiently informative that it appeared worthwhile to develop a program for the IBM 1620 to be employed in computing half-normal variates from multi-level factorial data. This program, Single Degree of Freedom Analysis of Variance (SIDOF), has a capacity of eight factors, each at two to ten levels. It requires as input the observations and normalized vectors of contrast coefficients α_p, β_q, γ_r, etc., where

FIGURE 2. HALF-NORMAL PLOT, SIX LARGEST CONTRASTS OMITTED



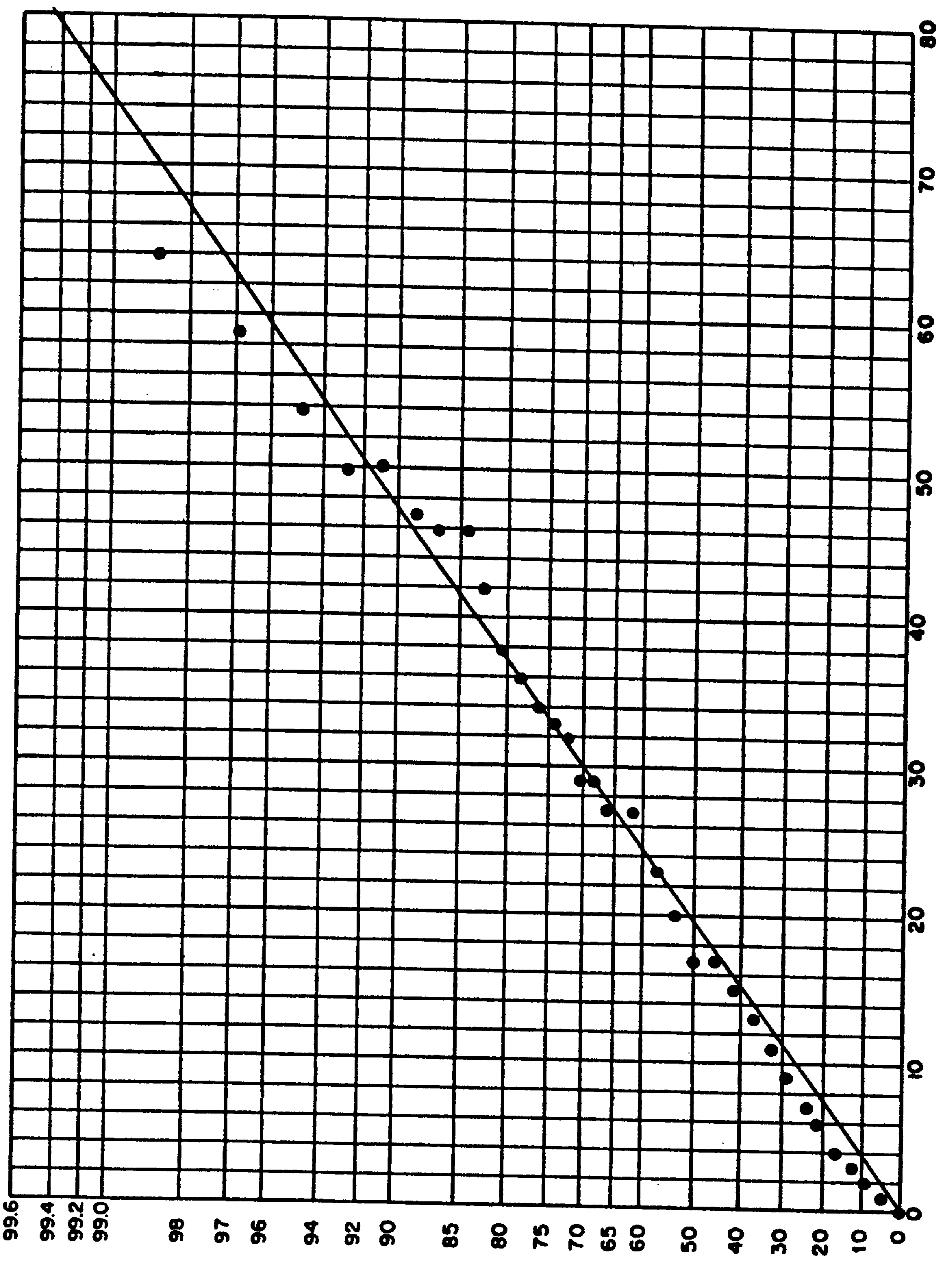


FIGURE 3. HALF-NORMAL PLOT, TEN LARGEST CONTRASTS OMITTED

$$\alpha_{ph} = a_{ph} / \sqrt{\sum_h a_{ph}^2}, \qquad h = 1, 2, \ldots, a;$$

$$\beta_{qi} = b_{qi} / \sqrt{\sum_i b_{qi}^2}, \qquad i = 1, 2, \ldots, b;$$

$$\gamma_{rj} = c_{rj} / \sqrt{\sum_j c_{rj}^2}, \qquad j = 1, 2, \ldots, c;$$

etc.

Each factor requires an additional "pass" through the machine. On the first pass, the machine computes the quantities (assuming three factors),

$$(A_p)_{ij} = \sum_h \alpha_{ph} \, y_{hij}.$$

On the second pass are computed the quantities.

$$(A_p B_q)_j = \sum_i \beta_{qi} \, (A_p)_{ij}$$

and on the third pass the quantities

$$(A_p B_q C_r) = \sum_j \gamma_{rj} \, (A_p B_q)_j.$$

At each pass the output includes both the (signed) contrasts developed and their squares. This program was one of the first developed for the IBM 1620 at Dugway Proving Ground and consequently was employed for a short period of time as a general-purpose analysis of variance. (It is, of course, much slower than other general-purpose programs available.)

6. EXPERIENCE. Some general comments on our experiences with half-normal plots for multi-level factorials may be in order. We shall be guided in this commentary largely by the approach of Daniel [1959].

a. Graph Sheets. We have generally used half-sheets of the Probability Scale x 90 Divisions paper available from Keuffel and Esser (Nos. 358-23 and 359-23).* Similar papers are available from several other sources. These papers are not particularly well-suited to the purpose. It would appear that special half-normal paper might be commercially feasible, but it is not, to our knowledge, currently available.

b. Birnbaum's test statistic. The test statistic developed by Allan Birnbaum [1959] has been used for our purposes. Birnbaum's work has been particularly oriented toward 2^p experiments and studies of the behavior of this statistic in multi-level factorials would be useful.

c. Defective values. Daniel indicates the utility of half-normal plotting in 2^p experiments for detecting defective values. For multi-level factorials the presence of defective values appears more difficult to diagnose, particularly with unrestricted sets of orthogonal contrasts. The isolation of the particular defective values is also more difficult.

d. Plot-splitting. The effect of plot-splitting upon the half-normal plots for multi-level experiments is similar to that described by Daniel. We have some reservations concerning indiscriminate searches for plot-splitting, however. It is generally accepted that in most experiments two-factor interactions tend to be smaller than main effects, three-factor interactions tend to be smaller than two-factor interactions, etc. (Here we are speaking of real effects and interactions, though perhaps of negligible magnitude.) Thus in actual experiments the slope of half-normal plots may be expected to increase with the relative number of high order interactions included. The plotted results of an experiment involving a number of small but real interactions may appear very similar to the results induced by plot-splitting, since split plot error contrasts invariably contain a relatively larger number of the higher order

*The graph sheets used in Figures 1, 2, and 3 were reproduced from a master kindly provided by Mr. Daniel. It is hoped that such sheets will soon be published.

contrasts. Our practice is generally to employ a split plot analysis only when knowledge of the experimental techniques indicates its propriety.

e. Convexity of plots. The detection of antilognormal distribution of error by downward convexity of half-normal plots appears difficult, as indicated by Daniel [1959, p. 336]. Most of our analysis work is, however, based on transformed data and we have seldom experienced this particular anomaly. In any event, the averaging effect of the contrasts would presumably minimize the effects of non-normality of error. On the other hand, we have noted that the removal of a moderate number of points representing apparently real effects often results in a downward convexity of the upper portion of the plot. We generally attribute this appearance to the inadvertent removal of one or more points representing error contrasts, for the result looks very much like the plot of a normal distribution with truncated upper tail.

7. REFERENCES.

Birnbaum, A. [1959], On the analysis of factorial experiments without replication, Technometrics I, No. 4, 343-57.

Davies, O. L., ed. [1954], The Design and Analysis of Industrial Experiments, Oliver and Boyd, London and Edinburgh.

Daniel, C. [1956], Fractional replication in industrial research, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, 87-98, Univ. Of California Press, Berkeley and Los Angeles.

Daniel, C. [1959], Use of half-normal plots in interpreting factorial two-level experiments, Technometrics I, No. 4, 311-41.

Daniel, C. [1962], Personal Communication.

PROPORTIONAL FREQUENCY DESIGNS

Sidney Addelman
Research Triangle Institute

CONDITION OF EQUAL FREQUENCIES. In 1945 Finney [5] introduced the procedure, known as fractional replication, which permitted the uncorrelated estimation of some of the effects and interactions when only a fraction of the full factorial arrangement was used. The standard method of constructing fractional replicate plans is to first choose an identity relationship and then deduce from this relationship the appropriate treatment combinations. By utilizing the assumption that the higher order interaction effects are negligible this standard procedure permits the estimation of the remaining effects. For the symmetrical factorial structure (all factors having the same number of levels) the standard procedure yields uncorrelated estimates due to the condition of equal frequencies of the factor levels. If the treatment combinations of the 2^5 factorial plan were inspected one would find that

(1) Each level of every factor occurs exactly eight times with every level of any other factor.

(2) Each combination of levels of any factor occurs exactly four times with every combination of levels of each pair of factors.

(3) Each combination of levels of any pair of factors occurs exactly two times with every combination of levels of any other pair of factors.

(4) Each level of any factor occurs exactly two times with every combination of levels of any three factors.

(5) Each level of any factor occurs exactly once with each combination of levels of any four factors.

(6) Each combination of levels of any pair of factors occurs exactly once with every combination of levels of any three factors.

Because the condition of equal frequencies is satisfied for all six of the above cases uncorrelated estimates of all effects can be obtained.

Now consider a 1/4 replicate of the 2^5 factorial structure defined by the identity relationship

$$I = ADE = BCD = ABCE$$

and consisting of the following treatment combinations:

A	B	C	D	E
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	0	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	0	1

One can verify that in this plan each level of any factor occurs exactly twice with every level of any other factor, and hence uncorrelated estimates of all main effects are obtainable, if all interactions are negligible. It can also be verified that each level of a factor does not occur the same number of times with every combination of levels of those pairs of factors with which it is aliased. Hence not all main effect estimates are uncorrelated with two-factor interaction estimates.

When one wishes to construct fractional replicate plans for symmetrical factorial arrangements one need only satisfy the appropriate equal frequency conditions to obtain uncorrelated estimates of the effects. However, in the construction of fractional replicate plans for asymmetrical factorial arrangements (all factors not having the same number of levels) the condition of equal frequencies requires more treatment combinations than are necessary to yield uncorrelated estimates.

CONDITION OF PROPORTIONAL FREQUENCIES. Although the equal frequency condition is sufficient to guarantee orthogonality of factors it is not a necessary condition. It was proved by Addelman and Kempthorne [3] that a necessary and sufficient condition that the main effect estimates of two factors be uncorrelated is that the levels of one factor occur with each of the levels of the other factor with proportional frequencies. Consider two factors, A and B, occurring at r and s levels respectively. Let

N = number of treatment combinations in the plan

$n_{i.}$ = number of times the i level of factor A occurs

$n_{\cdot j}$ = number of times the j level of factor B occurs

n_{ij} = number of treatment combinations in which the i level of factor A occurs with the j level of factor B.

The above necessary and sufficient condition for orthogonality can be displayed mathematically as

$$(1) \qquad n_{ij} = n_{i\cdot}n_{\cdot j}/N \quad .$$

The condition of proportional frequencies can be generalized so that plans may be constructed with permit uncorrelated estimates of two-factor interactions as well as main effects. Consider three factors A, B and C. In order that the interaction AB can be uncorrelated with C, each combination of the levels of A and B must occur with the levels of C with proportional frequencies, that is

$$(2) \qquad n_{ijk} = n_{ij\cdot}n_{\cdot\cdot k}/N \ .$$

Since it is desirable that A be uncorrelated with B

$$(3) \qquad n_{ij\cdot} = n_{i\cdot\cdot}n_{\cdot j\cdot}/N$$

and hence

$$(4) \qquad n_{ijk} = n_{i\cdot\cdot}n_{\cdot j\cdot}n_{\cdot\cdot k}/N^2 \quad .$$

This condition which assures that AB is uncorrelated with C also implies that AC is uncorrelated with B, BC uncorrelated with A, and hence AB, AC and BC are pairwise uncorrelated. If a plan contains four or more factors condition (4) must be replaced by

$$(5) \qquad n_{ijkm} = n_{i\cdot\cdot\cdot}n_{\cdot j\cdot\cdot}n_{\cdot\cdot k\cdot}n_{\cdot\cdot\cdot m}/N^3$$

which is the necessary and sufffficient condition that a plan permit uncorrelated estimation of all main effects and two-factor interaction effects.

COLLAPSING OF LEVELS. A factor at s_1 levels may be collapsed to a factor at $s_2 < s_1$ levels by making a many-one correspondence of the set of s_1 levels to the set of s_2 levels. If $s_1 = s_2^m$ then the s_1 levels can be collapsed to $(s_1-I)/(s_2-1)$ factors each having s_2 levels. Some illustrations of typical correspondence schemes are presented below.

Three-level factor		Two-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	0

Four-level factor		Three-level factor		Two-level factors
0	$\longrightarrow$	0	$\longrightarrow$	000
1	$\longrightarrow$	1	$\longrightarrow$	011
2	$\longrightarrow$	2	$\longrightarrow$	101
3	$\longrightarrow$	1	$\longrightarrow$	110

Five-level factor		Four-level factor		Three-level factor		Two-level factor
0	$\longrightarrow$	0	$\longrightarrow$	0	$\longrightarrow$	0
1	$\longrightarrow$	1	$\longrightarrow$	1	$\longrightarrow$	1
2	$\longrightarrow$	2	$\longrightarrow$	2	$\longrightarrow$	1
3	$\longrightarrow$	3	$\longrightarrow$	2	$\longrightarrow$	1
4	$\longrightarrow$	0	$\longrightarrow$	0	$\longrightarrow$	0

An orthogonal main-effect plan for the $2^2 \times 3^2$ experiment which permits uncorrelated estimates of all main effects with only nine treatment combinations is now constructed to illustrate the technique of collapsing levels. First construct an orthogonal main-effect plan for four factors, each having three levels with nine treatment combinations, namely

0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

If each of the first two factors are collapsed to two-level factors, the resulting treatment combinations constitute an orthogonal main-effect plan for the $2^2 \times 3^2$ experiment and are displayed below.

0	0	0	0
0	1	1	2
0	0	2	1
1	0	1	1
1	1	2	0
1	0	0	2
0	0	2	2
0	1	0	1
0	0	1	0

The smallest plan which yields uncorrelated estimates of the main effects of the $2^2 \times 3^2$ experiment and which also satisfies the equal frequency condition would require 36 treatment combinations.

It should be mentioned that the proportional frequency condition will be satisfied no matter what type of correspondence scheme is used to perform the collapsing procedure. However, the efficiency of the estimates depends upon the particular correspondence scheme chosen.

If the $(s_i - 1)$ degrees of freedom for each of the t_i factors at s_i levels are represented by $(s_i - 1)$ orthogonal contrasts among the s_i levels, the estimates obtained by these contrasts will be uncorrelated with the estimates obtained with the contrasts for any other factor, be-

cause the correspondence scheme automatically guarantees proportional frequencies of the levels of each factor.

REPLACING FACTORS. The collapsing procedure given above can be reversed so that a factor at s^m levels can replace $(s^m - 1)/(s - 1)$ factors, each at s levels. The replacement procedure can be illustrated by the construction of an orthogonal main-effect plan for the 3×2^4 experiment with eight trials. First construct an orthogonal main-effect plan for the 2^7 experiment with eight trials. The seven two-level factors can be represented by X_1, X_2, X_1X_2, X_3, X_1X_3, X_2X_3 and $X_1X_2X_3$. The treatment combinations for this plan are

0	0	0	0	0	0	0
0	0	0	1	1	1	1
0	1	1	0	0	1	1
0	1	1	1	1	0	0
1	0	1	0	1	0	1
1	0	1	1	0	1	0
1	1	0	0	1	1	0
1	1	0	1	0	0	1

It is known that there exists an orthogonal main-effect plan for the 2^3 experiment with four trials. The treatment combinations for this plan are 000, 011, 101, and 110. Thus, by choosing three factors of the 2^7 plan whose X representations are such that the generalized interaction of any two of the three factors is the third factor, three two-level factors can be replaced by a four-level factor, according to the following correspondence scheme:

Two-level factor				Four-level factor
0	0	0	$\longrightarrow$	0
0	1	1	$\longrightarrow$	1
1	0	1	$\longrightarrow$	2
1	1	0	$\longrightarrow$	3

Since the X representations of the first three factors of the above plan are X_1, X_2 and X_1X_2, these three factors can be replaced by a four-level factor and the orthogonal main-effect plan for the 4×2^4 experiment in eight trials is given by the following treatment combinations:

0	0	0	0	0
0	1	1	1	1
1	0	0	1	1
1	1	1	0	0
2	0	1	0	1
2	1	0	1	0
3	0	1	1	0
3	1	0	0	1

The plan for the 3×2^4 experiment is then obtained by collapsing the four-level factor to a three-level factor by the correspondence

Four-level factor		Three-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	2
3	$\longrightarrow$	1

The smallest plan which yields uncorrelated estimates of main effects in the 3×2^4 experiment and which also satisfies the equal frequency condition would require 24 treatment combinations.

The procedure for constructing plans which permit uncorrelated estimates of all main effects and some or all of the two-factor interaction

effects for asymmetrical factorial arrangements consists of first constructing the corresponding plan for a symmetrical factorial arrangement and then utilizing the collapsing or replacing techniques to obtain the desired plan. Whereas a plan permitting uncorrelated estimates of all main effects and all two-factor interactions among the two-level factors in the $2^3 \times 3^4$ experiment would require 72 treatment combinations to satisfy the condition of equal frequencies it would only require 27 treatment combinations to satisfy the proportional frequency condition.

BLOCKING. Even though the proportional frequency designs are highly fractionated they may still require more trials than can be carried out under uniform conditions. Thus, it would be desirable to divide the experimental data into smaller blocks in such a manner that the main effects may still be estimated without correlation. In order to perform an experiment in blocks one may utilize one or more of the factors of an orthogonal main-effect plan for the $4 \times 3^2 \times 2^6$ experiment with sixteen trials. The following plans may be derived from this one by using various factors as blocking factors:

(i) $4 \times 3^2 \times 2^5$ in 2 blocks of 8 treatment combiantions,

(ii) $4 \times 3^2 \times 2^3$ in 4 blocks of 4 treatment combinations,

(iii) $3^2 \times 2^6$ in 4 blocks of 4 treatment combinations,

(iv) $4 \times 3 \times 2^6$ in 4 blocks of 4 treatment combinations.

ORTHOGONAL POLYNOMIALS. The orthogonal contrasts which define effects and interactions in an equal frequency design can be readily determined from a table of orthogonal polynomials. The advantage of using orthogonal contrasts to define effects and interactions arises from the fact that orthogonal polynomials are so constructed that any term of the polynomial is independent of any other term. This property of independence permits one to compute each regression coefficient independently of the others and also facilitates testing the significance of each coefficient.

Tables of orthogonal polynomials for the case of equally spaced levels are readily available, e.g. Fisher and Yates [6], Anderson and Houseman [4]. It would be an impossible task to compile a general table of orthogonal polynomials for unequally spaced levels. However a simple procedure for computing these orthogonal polynomials is available and will be presented

below. If equally spaced levels do not each occur in a plan an equal number of times the published tables of orthogonal polynomials are not appropriate. The orthogonal polynomials for equally spaced levels which do not occur in a plan with equal frequency can be computed by the following method for unequally spaced levels.

For any set of orthogonal polynomials the linear contrast is of the form $\Sigma(\alpha + \beta x)y_x$, where α and β are constants, x is the level at which the factor occurs, y_x is the response to the treatment combination with the factor at the x level and the summation is over every value of x which is presented. The quadratic and cubic contrasts are of the form $\Sigma(\alpha + \beta x + \gamma x^2)y_x$ and $\Sigma(\alpha + \beta x + \gamma x^2 + \delta x^3)y_x$, respectively. The extension to higher order contrasts is obvious. Two contrasts are orthogonal if the coefficients of each contrast sum to zero and the sum of products of the corresponding coefficients of the two contrasts is zero.

Table 1

Coefficients of Orthogonal Contrasts

Level of x	Linear	Quadratic	Cubic
0	α	α	α
1	$\alpha + \beta$	$\alpha + \beta + \gamma$	$+\beta + \gamma + \delta$
2	$\alpha + 2\beta$	$\alpha + 2\beta + 4\gamma$	$+ 2\beta + 4\gamma + 8\delta$
4	$\alpha + 4\beta$	$\alpha + 4\beta + 16\gamma$	$+ 4\beta + 16\gamma + 64\delta$

We will illustrate the procedure for obtaining orthogonal polynomials for unequally spaced levels with an example.

Consider an independent variable x with levels 0, 1, 2 and 4. The coefficients of the linear, quadratic and cubic contrasts for this example are displayed in Table 1. The coefficients of the linear contrast must sum to zero. Thus,

$$4\alpha + 7\beta = 0.$$

Setting $\beta = 1$ we find that $\alpha = -7/4$. In order that the coefficients of the orthogonal contrasts be integers reduced to lowest terms we multiply these coefficients by 4 to obtain $\beta = 4$ and $\alpha = -7$. Substituting $\alpha = -7$ and $\beta = 4$ in the linear contrast given in Table 1, gives the linear coefficients.

Level of x	Coefficient of linear contrast
0	-7
1	-3
2	1
4	9

The coefficients of the quadratic contrast must sum to zero. Hence,

$$4\alpha + 7\beta + 21\gamma = 0$$

The sum of products of the corresponding coefficients of the linear and quadratic contrasts must also equal zero. Thus,

$$35\beta + 145\gamma = 0$$

Solving these two equations to obtain integral values for α, β and γ we obtain $\alpha = 14$, $\beta = -29$ and $\gamma = 7$.

If we substitute these values in the quadratic contrast and reduce the resulting coefficients to lowest terms the coefficients of the quadratic contrast is given by

Level of x	Coefficient of Quadratic contrast
0	7
1	-4
2	-8
4	5

Similarly the sum of the coefficients of the cubic contrast and the sum of products of the corresponding coefficients of the linear and cubic contrasts must each equal zero. Hence,

$$\begin{aligned} 4\alpha + 7\beta + 21\gamma + 73\delta &= 0 \\ 35\beta + 145\gamma + 581\delta &= 0 \\ 44\gamma + 252\delta &= 0 \end{aligned}$$

Solving these equations to obtain integral values for α, β, γ and we obtain $\alpha = -36$, $\beta = 392$, $\gamma = -315$ and $\delta = 55$. If we substitute these

values in the form of the coefficients of the cubic contrast given in Table 1 and reduce the resulting coefficients to lowest terms, the coefficients of the cubic contrast are given by

Level of x	Coefficients of Cubic contrast
0	-3
1	8
2	-6
4	1

The orthogonal polynomials are presented in the following table.

Table 2
Orthogonal Polynomials

Level of x	Linear	Quadratic	Cubic
0	-7	7	-3
1	-3	-4	8
2	1	-8	-6
4	9	5	1

The symbol β represents one unit of the linear effect of a factor when set equal to unity. In order to obtain integral coefficients β was set equal to 4 and hence $(1/4)\beta$ represents one unit of the linear effect. Consequently the linear contrast with coefficients given in Table 2 represents the estimate of $1/4$ the linear effect of the factor. It is easily verified that the coefficients of the quadratic contrast are given by

$$7 - \frac{29}{2} x + \frac{7}{2} x^2$$

where x = 0, 1, 2 and 4 respectively. Thus the symbol $\frac{2}{7}\gamma$ represents one unit of the quadratic effect, and the linear contrast with coefficients given in Table 2 represents the estimate of $\frac{2}{7}$ the quadratic effect of the factor.

Similarly it may be demonstrated that the cubic contrast with coefficients given in Table 2 represents the estimate of 12/55 the cubic effect of the factor.

This constant which is multiplying each effect will be denoted by $1/\lambda$ and in the tables of orthogonal polynomials presented by Addelman and Kempthorne [3] the value of λ and the sum of squares of the coefficients were both given. Any contrast defined by the coefficients given in the tables of orthogonal polynomials represents $1/\lambda$ times the appropriate effect of the factor.

EFFICIENCIES. Although any many-one correspondence of the set of s_1 levels to the set of s_i levels will yield proportional frequencies of the levels, there arises the problem of which correspondence is "best" in some sense. The problem may be solved by determining the efficiencies of the main-effect estimates obtained using proportional frequencies relative to the estimates which would result from using equal frequencies of the levels of each factor.

As an illustration we will calculate the relative efficiency of a three-level factor in a main-effect plan with twenty-five trials.

Assume the correspondence scheme used to collapse a five-level factor to three levels is as follows:

Five-level factor		Three-level factor
0	$\longrightarrow$	0
1	$\longrightarrow$	1
2	$\longrightarrow$	2
3	$\longrightarrow$	2
4	$\longrightarrow$	0

The levels 0, 1, and 2 occur in the ratio's 2: 1: 2. Thus for this factor the 0 level occurs in ten treatment combinations, the 1 level occurs in five treatment combiantions and the 2 level occurs in ten treatment combinations.

The variance of the linear effect estimate of this factor is equal to

$\sigma^2/20$ and hence the information on a unit basis is equal to $20/25\sigma^2 = 4/5\sigma^2$. The variance of the linear effect estimate of a three-level factor in 3^n trials is equal to $\sigma^2/2.\,3^{n-1}$ and the information on a unit basis is $2.\,3^{n-1}/3^n\sigma^2 = 2/3\sigma^2$. Hence the relative efficiency of the linear effect estimate is equal to $(4/5)x(3/2) = 6/5$.

The variance of the quadratic effect estimate for the three-level factor in twenty-five trials is equal to $\sigma^2/4$ and the information is then equal to $4/25\sigma^2$. The variance of the quadratic effect estimate with 3^n trials is equal to $\sigma^2/2.\,3^{n-2}$ and hence the information on a unit basis is equal to $2/9\sigma^2$. The relative efficiency of the quadratic effect estimate is therefore equal to $(4/25)x(9/2) = 18/25$.

Table 3

Relative Efficiencies of Proportional Frequency Estimates

Level	0 Proportional		1 frequency	Efficiency
	1	:	2	8/9
	2	:	3	24/25
	1	:	4	16/25
	3	:	4	48/49
	2	:	5	40/49
	1	:	6	24/49

Level Contrast	0		1 Proportional frequency		2	Efficiency
Linear	1	:	2	:	1	3/4
Quadratic	1	:	2	:	1	9/8
Linear	2	:	1	:	2	6/5
Quadratic	2	:	1	:	2	18/25
Linear	1	:	3	:	1	3/5
Quadratic	1	:	3	:	1	27/25
Linear	2	:	3	:	2	6/7
Quadratic	2	:	3	:	2	54/49
Linear	3	:	1	:	3	9/7
Quadratic	3	:	1	:	3	27/49
Linear	1	:	5	:	1	3/7
Quadratic	1	:	5	:	1	45/49

The relative efficiencies of the estimated effects are presented for various proportional frequencies in Table 3. One would choose the proportional frequencies which give the greatest efficiency of estimates. Thus for example, if an experiment in twenty-five trials involved two-level factors the two levels should occur in the ratio 2 : 3 rather than in the ratio 1 : 4 because the efficiency of the 2 : 3 ratio is 24/25 whereas the efficiency of the 1 : 4 ratio is only 16/25.

Bibliography

[1] Addelman, S. Orthogonal main-effect plans for asymmetrical factorial experiments. Technometrics, 4:21-46, 1962.

[2] Addelman S. Symmetrical and asymmetrical fractional factorial plans. Technometrics, 4:47-58, 1962.

[3] Addelman, S. and O. Kempthorne. Orthogonal main-effect plans. ARL 79, United States Air Force, November 1961.

[4] Anderson, R. L. and E. E. Houseman. Tables of orthogonal polynomial values extended to N=104. Agricultural Experiment Station Research Bulletin 297. Iowa State College, Ames, Iowa, 1942.

[5] Finney, D. J. The fractional replication of factorial arrangements. Annals of Eugenics, 12:291-301. 1945.

[6] Fisher, R. A. and F. Yates. Statistical tables for biological, agricultural and medical research. Oliver and Boyd, Edinburgh, 1938.

OPTIMAL DESIGN OF EXPERIMENTS

Herman Chernoff [1/]
Stanford University

1. INTRODUCTION. I would like to discuss some aspects of the theory of optimal design of experiments with particular emphasis on its relevance to the practice of statistics. There are two major branches of classical statistics, Estimation and Testing of Hypotheses, for which the theory of optimal design yields different results. Because of the time limitation, I shall confine my attention to certain results and examples in the theory of estimation.

2. SOME EXAMPLES. To illustrate the theory let us consider three examples. The first example is a well known one with a trivial solution. That is the one of estimating the slope of a regression (straight line). More specifically we have

Example 1.

The experimenter may choose any number y between -1 and $+1$. This number y designates an elementary experiment which corresponds to observing

$$Z = \alpha + \beta y + u$$

where u is normally distributed with mean 0 and variance 1 and α and β are unknown parameters. The experimenter is permitted to select a design consisting of n values $y_1, y_2, \ldots, y_n$, with possible repetitions. The design corresponds to performing the n designated experiments independently. It is desired to select a design which will yield the best possible estimate of the slope β.

It is well known and it is intuitively obvious that the best design consists of selecting $y = -1$ and $y = +1$ each half the time (providing n is even).

1/ This work was supported in part by Office of Naval Research Contract Nonr-225(52) at Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Another example which is of some current interest, having been discussed in yesterday's paper by Mr. Langlie [5] on a problem in reliability, and which is also relevant to the problem of Probit Analysis, may be expressed as follows:

Example 2.

A device, which may be used only once, can operate successfully under a stress s with probability

$$p = \int_{\frac{s-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad dt .$$

In other words one may say that the strength of the device, as measured by the maximum stress under which it will operate successfully, is normally distributed with unknown mean μ and variance σ^2. It is desired to select a design consisting of the choice of stress levels $s_1, s_2, \ldots, s_n$ which will yield an optimal estimate of $\mu - k\sigma$. The elementary experiment, designated s, consists of course of observing the success or failure of the device when used under stress s.

Finally a third problem which was discussed in detail in a recent paper of mine [2] deals with accelerated life testing. Here we wish to estimate the mean life time of a device when used under an environment of ordinary stress conditions. If, this mean lifetime is great and it is desired to have the estimate soon, then it is necessary to accelerate. The device is subjected to a much larger than ordinary stress. The results of such accelerated life testing can be relevant only if one assumes some form of relationship connecting the mean lifetime under various stresses. As an approximation we shall assume a quadratic relationship for some limited range. In addition since time is of the essence we shall assume that the cost of observing a device under stress s is proportional to the mean lifetime under that stress. Let us be more specific.

Example 3.

A device under stress environment s has lifetime T with an exponential distribution with failure rate (reciprocal of mean) given by

$$\varphi = \theta_1 s + \theta_2 s^2 \qquad \text{for } 0 \leq s \leq s^*$$

where θ_1 and θ_2 are unknown parameters. It is desired to estimate the failure rate under the ordinary stress s_o. This is

$$\varphi_o = \theta_1 s_o + \theta_2 s_o^2 .$$

An elementary experiment designated by s consists of observing the lifetime T of a device subjected to the environment s. The cost of the experiment s is

$$C(s) = c(\theta_1 s + \theta_2 s^2)^{-1} .$$

It is desired to select a design consisting of experiments $s_1, s_2, \ldots;$ $0 \leq s_i \leq s^*$, so as to obtain an optimal estimate of φ_o for a specified total cost.

Each of these examples has certain elements in common. Each may be regarded as a special case of the following general formulation. There is a set $\mathcal{E}$ of available elementary experiments e. In each case the distribution of the data of an experiment depends on the experiment and on k unknown parameters represented by $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. We wish to estimate some function $g(\theta_1, \theta_2, \ldots \theta_k)$ of the parameters. A design consists of the independent performance of experiments $e_1, e_2 \ldots$ with possible repetitions. It is desired to find a design which yields the best possible estimate of $g(\theta_1, \theta_2, \ldots, \theta_k)$ for a specified total cost or for a specified number of observations.

3. <u>THE LINEAR REGRESSION MODEL.</u> In 1952, Elfving [4] derived an elegant geometric solution to the optimal design problem for a special but important case of the above general formulation. As we shall see this result is applicable to a large variety of problems. Let ξ be a set of experiments e denoted by (y_1, y_2). The experiment e consists of observing

$$Z = \theta_1 y_1 + \theta_2 y_2 + u$$

where u is normally distributed with mean 0 and variance 1. It is desired to obtain an optimal estimate of $a_1 \theta_1 + a_2 \theta_2$ using a design consisting of n observations. The first example of estimating the slope of a straight line is a special case of Elfving's linear regression model where ξ is the set of points (1, y) with $-1 \leq y \leq 1$, and $(a_1, a_2) = (0, 1)$.

Elfving's solution consists of constructing a set S which is the smallest convex set containing the points (y_1, y_2) of ξ and their negatives $(-y_1, -y_2)$. Then extend the vector from (0, 0) to (a_1, a_2) until it penetrates the set S. The point of penetration (w_1, w_2) represents the optimal design. If this point is one of the original points (y_1, y_2) or $(-y_1, -y_2)$ the optimal design consists of repeating (y_1, y_2) n times. Otherwise the point of penetration is on a line segment connecting points corresponding to two of the original experiments (or their negatives). Then the optimal design consists of repeating these two experiments in proportions given by the distances from (w_1, w_2) to the two points. The greater proportion corresponds to the experiment closer to (w_1, w_2). Finally the variance of the least squares estimate based on this design is

$$\sigma_{\hat{\phi}}^2 = \left[n(w_1^2 + w_2^2)\right]^{-1}(a_1^2 + a_2^2) = a_1^2/nw_1^2 = a_2^2/nw_2^2$$

This solution can be illustrated with example 1. Here S is the square whose corners are (1, 1) and (-1, -1) corresponding to y = 1 and (1, -1) and (-1, 1) corresponding to y = -1. The line from (0,0) through $(a_1, a_2) = (0, 1)$ penetrates S at (0, 1) which is halfway between (1, 1) and (-1, 1). Thus the optimal design consists of repeating the experi-

ments corresponding to $y = 1$ and $y = -1$ each half the time (as was well known). Furthermore the variance of the estimate of β should be $1/n$.

Elfving's result applies in the obvious fashion to experiments involving k parameters. Here we need repeat at most k of the available experiments in certain proportions to obtain the optimal estimate.

4. RESULTS FOR THE MORE GENERAL PROBLEM. As mentioned in the preceeding section the problem treated by Elfving is a special case of the more general one formulated in section 2. For this more general problem, related results have been obtained [1]. These results concern designs which are asymptotically locally optimal. We shall defer the interpretation of these adjectives until the discussion of Example 2 in section 5.

It was shown that asymptotically locally optimal designs depend on the form of the matrix $J(e)$ which is defined as Fisher's information matrix divided by the cost of the experiment e. In other words if experiment e has cost $C(e)$ and yields data X with probability distribution $f(x, \theta, e)$, Fisher's information matrix is

$$I(e) = \left\| E\left\{ \frac{\partial \log f(X, \theta, e)}{\partial \theta_i} \frac{\partial \log f(X, \theta, e)}{\partial \theta_j} \right\} \right\|$$

and the information per unit cost is

$$J(e) = I(e)/C(e).$$

Clearly if the cost of experimentation is constant one need concern oneself only with $I(e)$. The relevance of Fisher's Information derives from its well known additive properties and the fact that the maximum-likelihood estimate $\hat{\theta}_n$, based on the outcome of n independent repetitions of e, has an approximately normal distribution with mean θ and covariance matrix $[nI(e)]^{-1}$ for large n.

When it is desired to estimate one function of the k parameters, there are asymptotically locally optimal designs which involve at most k of the experiments of ξ in certain proportions. This result which corresponds to one of Elfving's results, together with the use of Fisher's Information, permits one to reduce the calculation of optimal designs to the maximization of a function of a fixed number of variables.

In the linear regression problem of Elfving, the information matrix for $e = (y_1, y_2)$ is

$$I = \| y_i y_j \| = J.$$

Since asymptotically optimal designs are determined by the information per unit cost it follows that for any problem where $J(e)$ can be put in the above form, the solution is the same as Elfving's with a_i replaced by $\frac{\partial g}{\partial \theta_i}$.

The illustration of the next section will help clarify the meaning of these results. In the meantime it may be remarked that if for each experiment the distribution of the outcome depends on only one function of the parameters, $J(e)$ can be put in the above form and Elfving's results are applicable. In particular they are applicable to both examples 2 and 3.

5. ILLUSTRATION. We shall find it informative to illustrate the method with example 2. Here the outcome of the experiment s is success or failure where the probability of success is

$$p(s, \mu, \sigma) = \int_{\frac{s-\mu}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left(\frac{s-\mu}{\sigma}\right)$$

where Φ is the normal cdf. In other words the role of the density $f(X, \theta, e)$ is played by

$$f = p^X(1 - p)^{1-X}$$

where $X = 1$ for success and 0 for failure.

$$\log f = X \log p + (1-X) \log(1 - p)$$

$$\frac{\partial \log f}{\partial \mu} = \frac{X - p}{p(1 - p)} \frac{\partial p}{\partial \mu}$$

$$\frac{\partial \log f}{\partial \sigma} = \frac{X - p}{p(1 - p)} \frac{\partial p}{\partial \sigma} .$$

Since $E\{(X-p)^2\} = p(1-p)$,

$$J(s) = I(s) = \left[p(1 - p)\right]^{-1} \left\| \begin{matrix} \left(\frac{\partial p}{\partial \mu}\right)^2 & \frac{\partial p}{\partial \mu}\frac{\partial p}{\partial \sigma} \\ \frac{\partial p}{\partial \mu}\frac{\partial p}{\partial \sigma} & \left(\frac{\partial p}{\partial \sigma}\right)^2 \end{matrix} \right\|$$

$$J(s) = \left\| y_i y_j \right\|$$

where

$$y_1(s) = \left[p(1-p)\right]^{-1/2} \frac{\partial p}{\partial \mu} = \left[2 \pi p(1-p)\right]^{-1/2} \sigma^{-1} \exp\left[-(s-\mu)^2/2\sigma^2\right]$$

and

$$y_2(s) = \left[p(1-p)\right]^{-1/2} \frac{\partial p}{\partial \mu} = \left[2 \pi p(1-p)\right]^{-1/2} (s-\mu)\, \sigma^{-2} \exp\left[-(s-\mu)^2/2\sigma^2\right].$$

Next we plot the set of points $\left[y_1(s),\ y_2(s)\right]$ in Figure 1. We add the negatives of these points and construct S the smallest convex set containing them. We note that for $s = \mu + t\sigma$, $y_2(s)/y_1(s) = t$. We also note the curve of $\left[y_1(s),\ y_2(s)\right]$ reaches its maximum and minimum at $s = \mu \pm k_o \sigma$ where $k_o = 1.57$. Finally, since we wish to estimate $\mu - k\sigma$,

we draw the vector from (0, 0) through (1, -k), i.e. the line through the origin with slope -k, and note where it penetrates the convex set S.

Clearly there are two cases.

Case 1. $|k| \le k_o$. Here the vector penetrates S at one of the original $[y_1(s), y_2(s)]$ points. In fact this point corresponds to $s = \mu - k\sigma$ and hence the optimal design consists of using $s = \mu - k\sigma$ for all observations.

Case 2. $|k| > k_o$. Here the vector penetrates S at the straight line section of the boundary. The optimal design consists of applying the stress levels $\mu - k_o\sigma$ and $\mu + k_o\sigma$ in proportions $k+k_o$ to $k-k_o$. In cases 1 and 2 the formal application of the formula for the variance of the maximum likelihood estimate of $\mu - k\sigma$ based on the optimal design is given by

$$2\pi\sigma^2 \Phi(k)\left[1-\Phi(k)\right] \; e^{k^2} n^{-1}$$

in case 1, and

$$2\pi\sigma^2 \Phi(k_o)\left[1-\Phi(k_o)\right] \; e^{k_o^2} k_o^{-2} k^2 n^{-1} = 1.64\,\sigma^2 k^2 n^{-1}$$

in case 2.

6. THE RELEVANCE OF OPTIMAL DESIGN. Now we shall find the illustrative example helpful in interpreting the results of the theory of optimal design of experiments and in understanding its relevance in practical applications. For simplicity let us confine our attention to case 2 at first.

First we note one very peculiar aspect of the optimal design. Since it involves using stress levels $\mu - k_o\sigma$ and $\mu + k_o\sigma$, to apply it one must know μ and σ. But if one knew μ and σ, there would be no need to experiment. While this seems to be ridiculous, a glance at figure 1 indicates that if one used an approximation to $\mu \pm k_o\sigma$, one would have a rather good approximation to the optimal design. Thus there is surprisingly little loss of efficiency when one is not certain about μ and σ. It is this property that the word local is used to describe. In other words our design would be efficient if we knew the parameters and

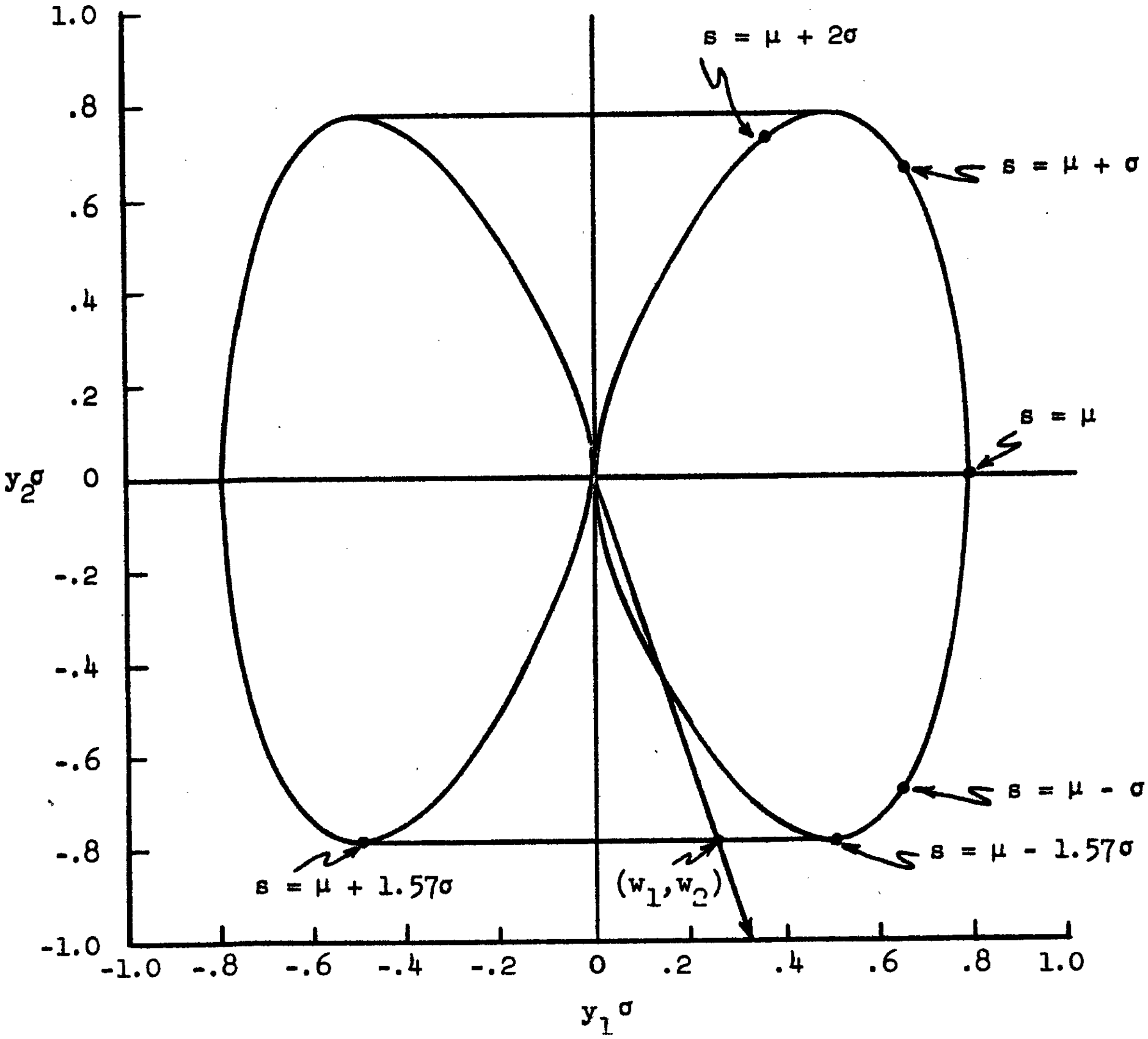

s = μ + 2σ
s = μ + σ
s = μ
s = μ - σ
s = μ - 1.57σ
s = μ + 1.57σ
(w1, w2)
1.0
.8
.6
.4
.2
0
-.2
-.4
-.6
-.8
-1.0
y2σ
-1.0
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1.0
y1σ

Figure 1

is approximately efficient if we use an approximation to the unknown parameters.

This raises the issue of the adjective <u>asymptotic</u>. If one had a <u>large</u> sample available, one could use some of the initial observations to derive an initial estimate of and on which to base an approximation to the optimal design. Furthermore the qualification asymptotic derives from a couple of other aspects. First, the properties relating the variance of the approximate distribution of the maximum likelihood to the information matrix and giving the efficiency of this estimate is based on asymptotic theory assuming large sample size. A second and relatively minor point, is illustrated by example 1 if an odd number of observations are available. The optimal design calls for putting half the observations at +1 and half at -1. This is impossible in a trivial way when n is odd. On the other hand the effect of this impossibility is negligible when n is large.

Having seen how we must qualify the tern optimal by the adjectives local and asymptotic, we can now consider a more fundamental issue. Briefly, our optimal design is simply impractical. Only in the rather unrealistic context where I had absolute faith in the model would I consider this as a solution. In fact, any reasonable statistician would insist on using several other stress levels at least to check on the model.

Another unreasonable aspect of our optimal design arises from its derivation based on the single minded purpose of obtaining a good estimate of <u>one</u> function $g(\theta_1, \theta_2, \ldots, \theta_k)$ of the parameters. In many practical problems, experimentation is used to serve several purposes simultaneously.

One may reasonably inquire about what function does the theory of optimal design serve, if (1) the optimality must be qualified as locally asymptotically optimal and (2) the designs it yields are unreasonable. Basically the functions are the following. First, the theory provides a yardstick for comparison purposes. If the designs proposed yesterday by Mr. Langlie, or the Up and Down Method [3, p. 319], or some other practical design turns out to be relatively efficient compared to our solution (as measured by asymptotic variance) then clearly there is no point in attempting to improve on this aspect of these methods. If, on the other hand, one of these methods were to have a low efficiency, then one is forced to delve deeper to see what, if anything, can be done to improve the design.

Second, theory not only presents an optimal design but indicates rather clearly how this design can be modified with relatively low loss of efficiency. The theory serves to direct the attention of the practical statistician toward designs which combine relatively high efficiency with practical utility when robustness and multi-purpose considerations are taken into account.

7. MISCELLANEOUS COMMENTS. I would like to conclude this paper with a few assorted comments. First, the proposed solution to example 2 in case 1 when $|k| \leq k_0$ consists of repeating one experiment n times. Not only is this solution impractical, but from a theoretical point of view it represents a degenerate situation. When a single level s is used, one can use the data to estimate only

$$p(s, \mu, \sigma) = \int_{\frac{s-\mu}{\sigma}}^{\infty} (2\pi)^{-1/2} e^{-t^2/2}\, dt$$

or functions of $p(s, \mu, \sigma)$. Then one can check whether $\frac{s-\mu}{\sigma}$ is in fact close to k (as it should be if the design were optimal). But not knowing σ, one can not estimate $\mu - k\sigma$. Thus the formula for the asymptotic variance presented at the end of section 5 is meaningful only as an approximation to the case where several levels of stress close to the optimal one were used. Alternatively one could regard $p(1-p)n^{-1}$ as the asymptotic variance of the estimate of p.

For a large sample sequential procedure, it seems clear that our theory is applicable. If one were to reestimate the parameters after each observation, and use these estimates to derive approximations to the optimal design, the resulting procedure should be asymptotically optimal in the sequential version and the adjective local need not be applied.

What is more interesting, perhaps, is the study of the "not so large" sample sequential case. Here even the following seemingly simple problem proposed by Harold Gumbel does not have a simple solution. Suppose that experiment e_i yields observation X_i which is normally distributed with unknown mean μ and unknown variance σ_i^2, $i=1,2$, and it is desired to estimate μ. In other words, two measuring instruments of unknown accuracy are available. How should one select between the

two experiments sequentially so as to obtain a good estimate efficiently when the sample size is not necessarily very large?

BIBLIOGRAPHY

(1) Chernoff, H. (1953), Locally optimal designs for estimating parameters, Ann. Math. Statist. 24 586-620.

(2) Chernoff, H. (1962), Optimal accelerated life designs for estimation, Technometrics 4, 381-408.

(3) Dixon, W. J., and Massey, F. J. (1957), Introduction to Statistical Analysis, (2nd ed.) McGraw Hill, New York.

(4) Elfving, G. (1952), Optimal allocation in linear regression theory, Ann. Math. Statist. 23, 255-262.

(5) Langlie, H. J. (1962), A Reliability Test Method for "One-Shot" Items; presented at the Eighth Conference on The Design of Experiments in Army Research, Development and Testing.

VIBRATION EXPERIMENTS

F. Pradko
Dynamic Simulations Laboratories
U. S. Army Ordnance Tank-Automotive Command

The design of military vehicles is a rather complicated mixture of many technical activities. In each new development of a tank, truck or jeep, a substantial amount of engineering ideas are required to be blended together to bring forward vehicles possessing features and merit of advanced capability.

While each vehicle development is a separate and distinct program, there are development goals and problems that continually reappear and appear to be common. For example, it is always important to create a good suspension system and it is equally important to provide a dynamically stable vehicle.

The Suspension System is vital for it determines the vehicle ride and vibration behavior and it also establishes the tolerable speed limit of travel over various terrain surfaces, both for the man and the machine.

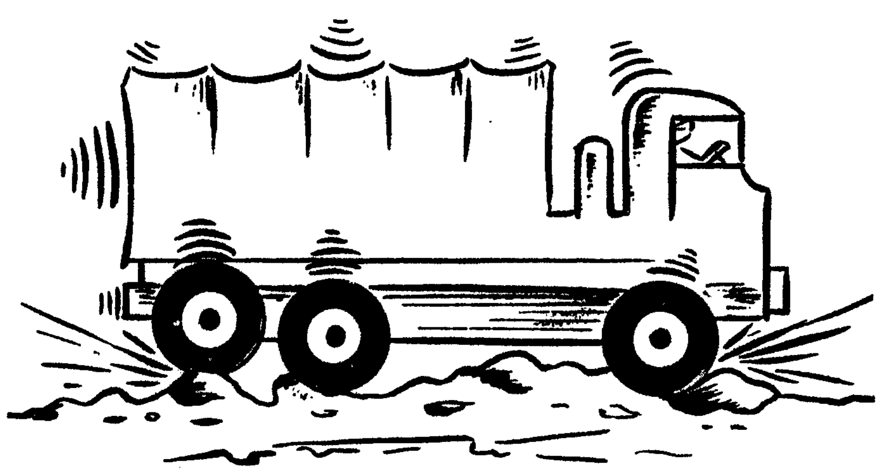


FIG. 1. STEP ONE

Vehicle stability is a design area that also receives considerable attention, particularly in combat vehicle programs where large caliber weapons are expected to be fired from chassis of reduced weight and decreased size. Within military circles reference to vehicle stability differs from the usual automotive connotation. In place of steering behavior or directional control, vehicle stability pertains to pitch and roll movement, resulting from the gun recoil forces.

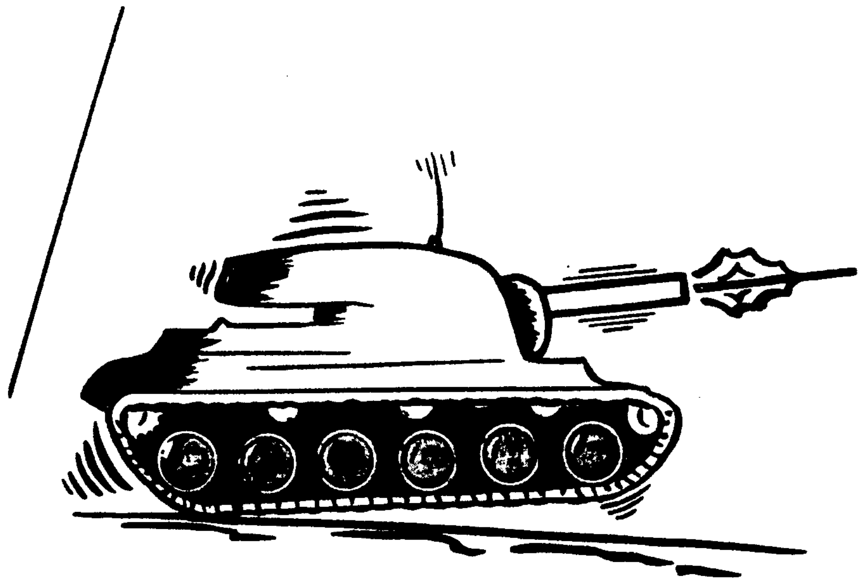


FIG. 2. STEP TWO

These two elements of vehicle design can be considered as perennials. They are always around and unfortunately they are rather "tough nuts" to handle. Individually they present major stumbling blocks to design engineers. Normally, useful study of these problems is beyond the level of developing a new layout or "cranking" through several equations on a desk calculator. As a result, the magnitude of each task has produced design specialists. These people, however, are not medicine men who consistently are able to generate successful answers.

They need help; they need either physical or analytical means to guide and measure their design approaches. Consequently, knowledge of these systems must be consistently bolstered and expanded. This demand requires unique capability:

First, since all wheeled and tracked vehicles are earth bound, knowledge of road profiles or terrain contours upon which they move must be secured.

Next, detailed mathematical models are necessary that describe the vehicle and how it reacts to external disturbances or internal design changes.

Then, an accurate recording procedure is essential to transmit results from high speed computers in such format that their meaning may be assessed graphically, visually or physically.

The benefits of integrating these steps would be a complete capability for realistic design evaluation comparable to controlled tests at a proving ground.

Accordingly, the Army Tank-Automotive Command sought a means to physically simulate the suspension performance and stability dynamics of a design while it was still in the blueprint stage, and to bring together the theory, mathematics and computer machinery to study each of these problems indoors in the laboratory. It was also the feeling that if this program was to be successful, it must produce the ability to predict behavior and allow practicing engineers the opportunity to pretest their designs before commitment to fabricate expensive wood mock-ups, experimental test rigs or engineering prototypes.

Suspension Simulation:

Simulation of vehicle suspension systems is basically the task of predicting the motion response of the vehicle to disturbances from the road. A thorough understanding of the elements that constitute this system and their interrelationship is essential to such study. Beginning with the road is perhaps logically the first step. The basic requirement is to present to the wheels, springs, and shock absorbers, the vertical displacement and frequency identical to those existing in road or terrain surfaces. For this purpose it is possible to construct

a computer model of a road profile using either digital or analog computer techniques. The basic data profile information may be secured by conventional rod and transit means or automatic measuring instruments.

In real life the road is a stationary wave form over which the vehicle travels. The relationship between car velocity and the static road generates the suspension dynamics.

In a computer simulation the vehicle's forward movement cannot be faithfully reproduced. To conveniently maintain an order of reality the road is moved instead. The road profile is presented to the suspension components, as a continuous rearward velocity that normally would be road speed.

To accurately duplicate the physical case the road is presented to each wheel separately, properly phased so that the rear wheels "see" the same road irregularities as the front wheels, although at a later time. This phasing is governed by the wheel base and vehicle speed.

One successful procedure of road profile generation utilizes a digital computer and a digital-to-analog converter. The digital machine stores the road profile data in elevation increments. It selects the elevation that each wheel requires at a particular time and generates the time between elevations.

Several preliminary considerations which must be resolved before the construction of such a digital road function include:

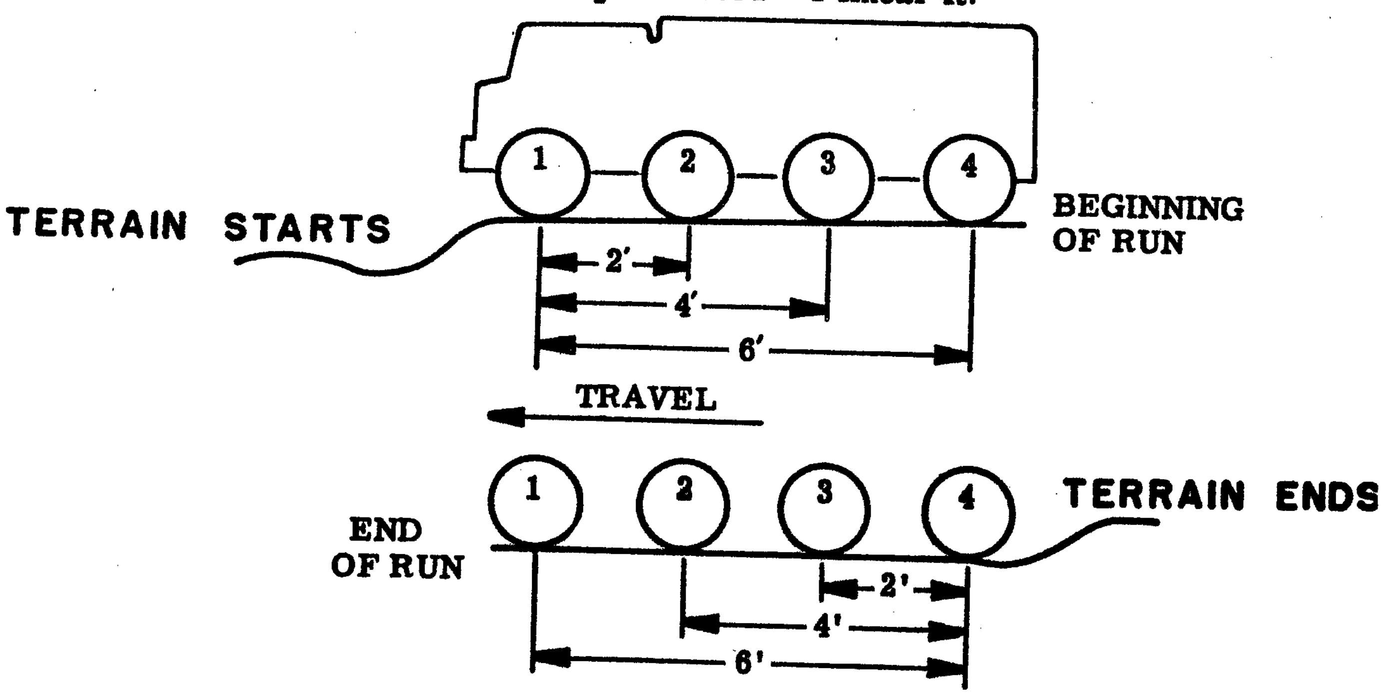
1. The number of vehicle wheels.
2. The spacing between wheels.
3. The starting road level.
4. The number of computer cells per road increment.
5. Overall length of the road profile.

The time at which a particular point on the road will arrive at each wheel is determined by the wheel spacing. This spacing is also considered in deciding how many elevation values will be equivalent to one linear foot of road. The following example will illustrate these points.

A vehicle suspension is set up on the analog computer; this simulation is for one side of the vehicle only, it being assumed that the other side is identical. The vehicle has four wheels on a side, spaced two feet apart. It will be driven over a Belgian Block type road. The road consists of 307 elevations spaced one foot apart. These things being known, it is possible to prepare a scheme for generating the road function which will pass under each wheel in sequence. To rerun the road after once traversing it, a starting road level must be assumed, usually the initial starting elevation, or very near to it. For the conditions just outlined a situation similar to that shown in Figure 3 will exist.

Preliminary assumptions:

1. Vehicle is sitting on road level.
2. Start of road strikes 1st wheel
3. One computer word = 1 linear ft.



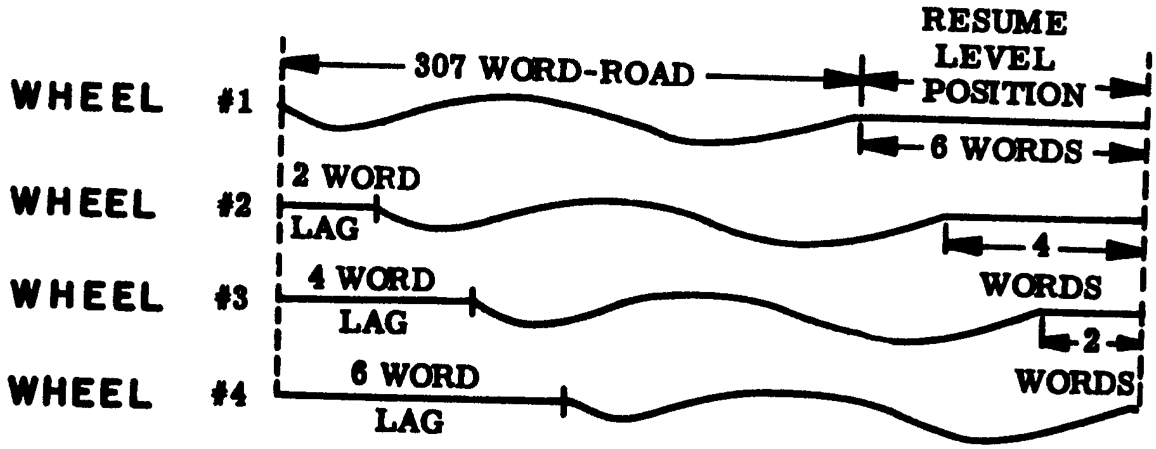


FIG. 3. ROAD PROFILE

The best procedure in this instance is to let one foot of road be represented by one computer word. However, if the spacing between wheels is uneven, a scheme utilizing several words to the linear foot would be required.

After the preliminary road function details have been accounted for, the actual generation of the road function can be undertaken. This naturally divides into the following steps:

1. Preparation of the data tape.

2. Placing road level data into computer memory.

3. Generation of the road function tape.

4. Transfer of the road function data to the digital-to-analog converter, etc.

Each computer "word" of information contains five channels, four of which are used for terrain simulation. Each of these

channels may represent data. An algebraic representation of such a word is 1 aa bb cc dd 00, where each pair of letters represents one channel in the output of the Digital-Analog conversion system, while the number (1) in the sign position designates a particular group of D-A Converters, (Note that since the last channel is not used it is represented by 00, i.e. no information present). With proper scaling and programming, each channel can become a road-profile-wheel-terrain-function-generator (RPWTFG). Hence, there are four RPWTFG's per computer word.

The time between data increments on the computer is generated by using a time control subroutine which increases or decreases the time between "calling up" the data increments. The speed of the road function is determined from the recorder tracings as follows:

$$\text{mph} = \frac{\text{actual length of road (ft)}}{\text{length of converted road (mm)}} \times \text{paper speed} \left(\frac{\text{mm}}{\text{sec.}}\right) \times \frac{30 \text{ mph}}{44 \text{ ft/sec}}$$

By this procedure vehicle road speed is established. Maximum road speed is only limited by the computers ability to call up data. Utilization of a time delay subroutine within the computer facilitates decreasing the speed. Also, greater speed can be attained by shortening the road, i.e. picking up every second or every third road elevation. Oscillograph recordings, as generated by this system, are illustrated in Figure 4.

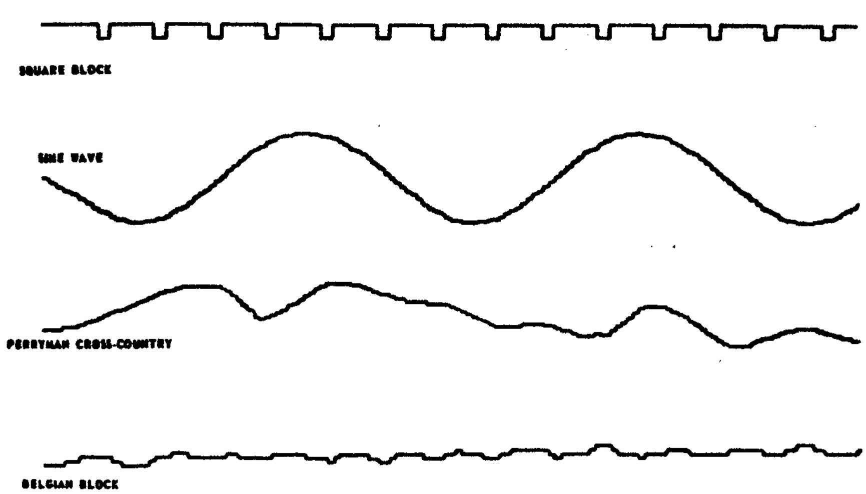


FIG. 4. ROAD PROFILE

A magnetic tape recorder and reproducer add convenience and efficiency to the system. By recording successive speeds on magnetic tape, the digital computer is used only once for a particular vehicle. In addition, velocity multiplication may be obtained by recording at one speed and reproducing at higher speeds.

Thus, any terrain that can be numerically described as increments of elevation with respect to horizontal distance, can be simulated with a digital computer for the analysis of vehicle behavior.

With the road prepared the simulation requires a model of the suspension system. For this purpose the Analog computer is best suited. The computer, as used, provides an accurate representation of the design. In a true sense the computer is an electronic model of the vehicle. The degree of realism achieved is naturally governed by the quantity of vehicle characteristics simulated.

As in any simulation, the system is first described by a mathematical model. The equations represent the dynamic system - the vehicle chassis, the suspension system and the road surface input.

Vehicular vibration components include the mass and inertia of the sprung components, the suspension springs, shock absorbers, road wheel masses, and the spring and damping characteristics of the wheel assembly.

The typical method of describing a vehicle to be simulated is shown in Figure 5 and Figure 6. From the diagrams, the equations of motion may be stated. These expressions are written as common linear differential equations with non-linear coefficients.

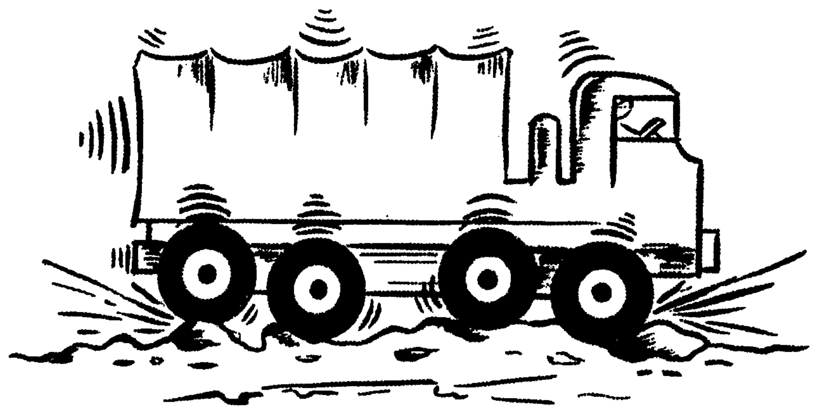


FIG. 5. VEHICLE CONFIGURATION

Non-linearities in the simulation exist normally due to non-linear spring characteristics, double acting shock absorbers, and the fact that wheels may leave the road surface.

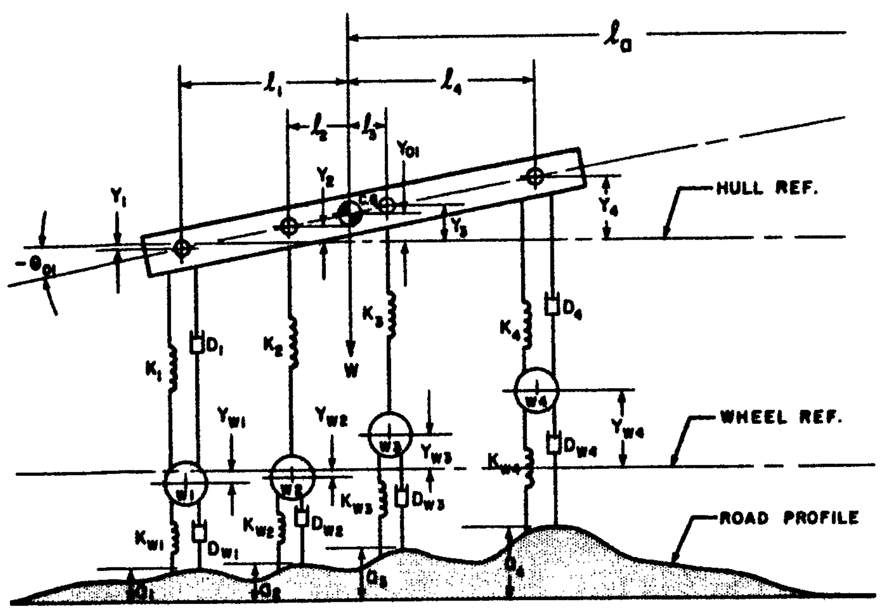


FIG. 6. VEHICLE SCHEMATIC

The equations result from linear and angular counterparts of Newton's second law of motion. The summation of the vertical forces on the chassis equals the mass of the chassis times its vertical acceleration; the summation of the torques about the center of gravity is equal to the polar moment of inertia times the angular acceleration of the hull. The forces and torques result from relative displacement and velocity of the springs and shock absorbers respectively. For example, the vertical force of the front wheel spring is equal to the spring constant (K_1) times the relative displacement between the wheel and the chassis just above the wheel. Similarly, the torque is a product of this force times the distance from the center of gravity. The shock absorber force is a product of the damping coefficient and the relative velocity between the wheel and the chassis above the wheel.

Chassis, pitch and bounce equations may be developed using these relationships. Separate differential equations are written to describe the motion of each wheel. Auxiliary equations are written to relate the pitch motion to the vertical motion so that the displacement and velocity of the chassis at each wheel station may be found.

Each wheel of the vehicle is considered a separate mass, spring, and damper system connected to the ground and to the chassis, the link to the chassis being the suspension spring and shock absorber. The force exerted on the wheel by the ground is equal to a product of the wheel rubber displacement and spring constant. This force may have only one sign since the ground cannot "pull" down on the wheel. The suspension spring force is also exerted on the wheel.

Simulation Equations:

Chassis Vertical Motion:

$$\ddot{Y}_o = \frac{\sum F_y}{M_o} \quad \text{(c.g. Bounce Acceleration)}$$

$$\ddot{Y}_o = -\frac{K_1}{M_o}(Y_1 - Y_{w1}) - \frac{K_2}{M_o}(Y_2 - Y_{w2}) - \frac{K_3}{M_o}(Y_3 - Y_{w3})$$

$$-\frac{K_4}{M_o}(Y_4 - Y_{w4}) - \frac{D_1}{M_o}(\dot{Y}_1 - \dot{Y}_{w1}) - \frac{D_4}{M_o}(\dot{Y}_4 - \dot{Y}_{w4}) + g$$

Chassis Pitch Motion:

$$\ddot{\theta} = \frac{\sum T}{J_o} \quad \text{(c.g. Angular Acceleration)}$$

$$\ddot{\theta} = -\frac{K_1 l_1}{J_o}(Y_1 - Y_{w1}) - \frac{K_2 l_2}{J_o}(Y_2 - Y_{w2})$$

$$+ \frac{K_3 l_3}{J_o}(Y_3 - Y_{w3}) + \frac{K_4 l_4}{J_o}(Y_4 - Y_{w4})$$

$$- \frac{D_1 l_1}{J_o}(\dot{Y}_1 - \dot{Y}_{w1}) + \frac{D_4 l_4}{J_o}(\dot{Y}_4 - \dot{Y}_{w4})$$

Vertical Wheel Motion:

$$\ddot{Y}_w = \frac{\sum F_w}{M_w}$$

$$\ddot{Y}_{w1} = -\frac{K_{w1}}{M_{w1}}(Y_{w1} - a_1) - \frac{D_{w1}}{M_{w1}}(\dot{Y}_{w1} - \dot{a}_1) + \frac{K_1}{M_{w1}}(Y_1 - Y_{w1})$$

$$+ \frac{D_1}{M_{w1}}(\dot{Y}_1 - \dot{Y}_{w1}) + g$$

$$\ddot{Y}_{w2} = -\frac{K_{w2}}{M_{w2}}(Y_{w2} - a_2) - \frac{D_{w2}}{M_{w2}}(\dot{Y}_{w2} - \dot{a}_2) + \frac{K_2}{M_{w2}}(Y_2 - Y_{w2}) + g$$

$$\ddot{Y}_{w3} = -\frac{K_{w3}}{M_{w3}}(Y_{w3} - a_3) - \frac{D_{w3}}{M_{w3}}(\dot{Y}_{w3} - \dot{a}_3) + \frac{K_3}{M_{w3}}(Y_3 - Y_{w3}) + g$$

$$\ddot{Y}_{w4} = -\frac{K_{w4}}{M_{w4}}(Y_{w4} - a_4) - \frac{D_{w4}}{M_{w4}}(\dot{Y}_{w4} - \dot{a}_4) + \frac{K_4}{M_{w4}}(Y_4 - Y_{w4})$$

$$+ \frac{D_4}{M_{w4}}(\dot{Y}_4 - \dot{Y}_{w4}) + g$$

Auxiliary Chassis Equations:

$$Y_{1-4} = Y_o + \ell_{1-4} \,\text{SIN}\, \theta \qquad \dot{Y}_{1-4} = \dot{Y}_o + \ell_{1-4} \,\text{SIN}\, \theta$$

The non-linearities are best described in graphical form, as is shown in Figures 7 and 8.

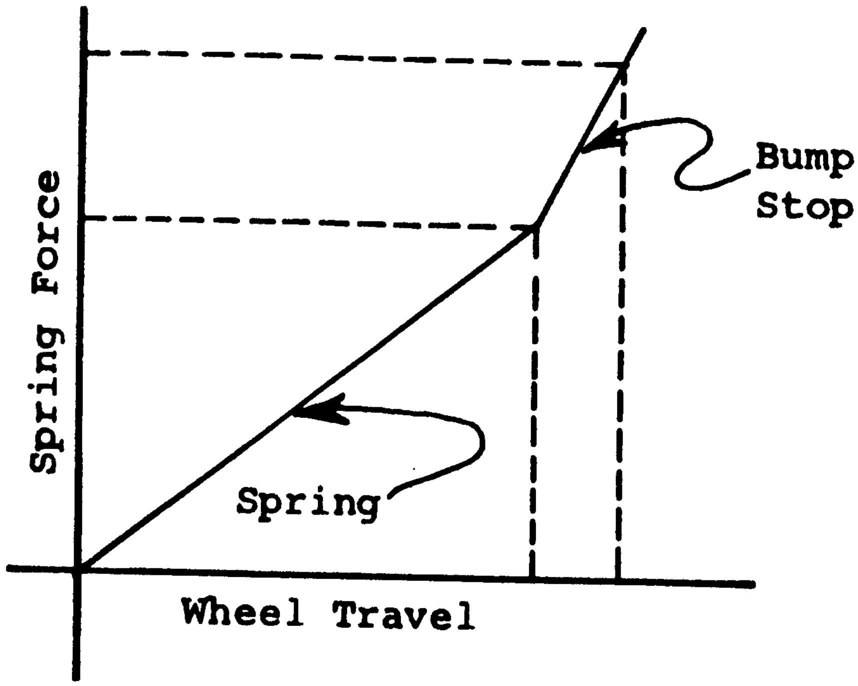


FIG. 7. SPRING LOAD VS WHEEL DEFLECTION

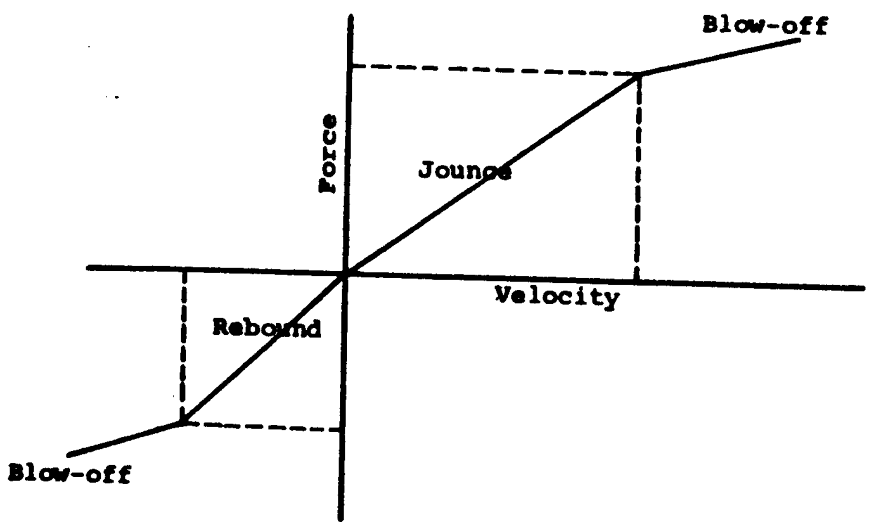


FIG. 8. SHOCK ABSORBER VS VELOCITY AT WHEEL

The non-linear suspension springing is composed of two linear segments, the one of lesser slope being the suspension spring, and the other the bump stop. The shock absorber non-linearity is shown in Fig. 8, which has four linear segments simulating different rates in compression and expansion with blow-off valves. The circuitry for creating the significant segments of a suspension simulation are shown in Figures 9 - 13. If a wheel leaves the ground, no spring force can exist between the ground and the wheel. To provide for this realistic action, a diode representing a unidirectional spring force is put in series with the wheel feedback loop, as is shown in the wheel circuit, Figure 11.

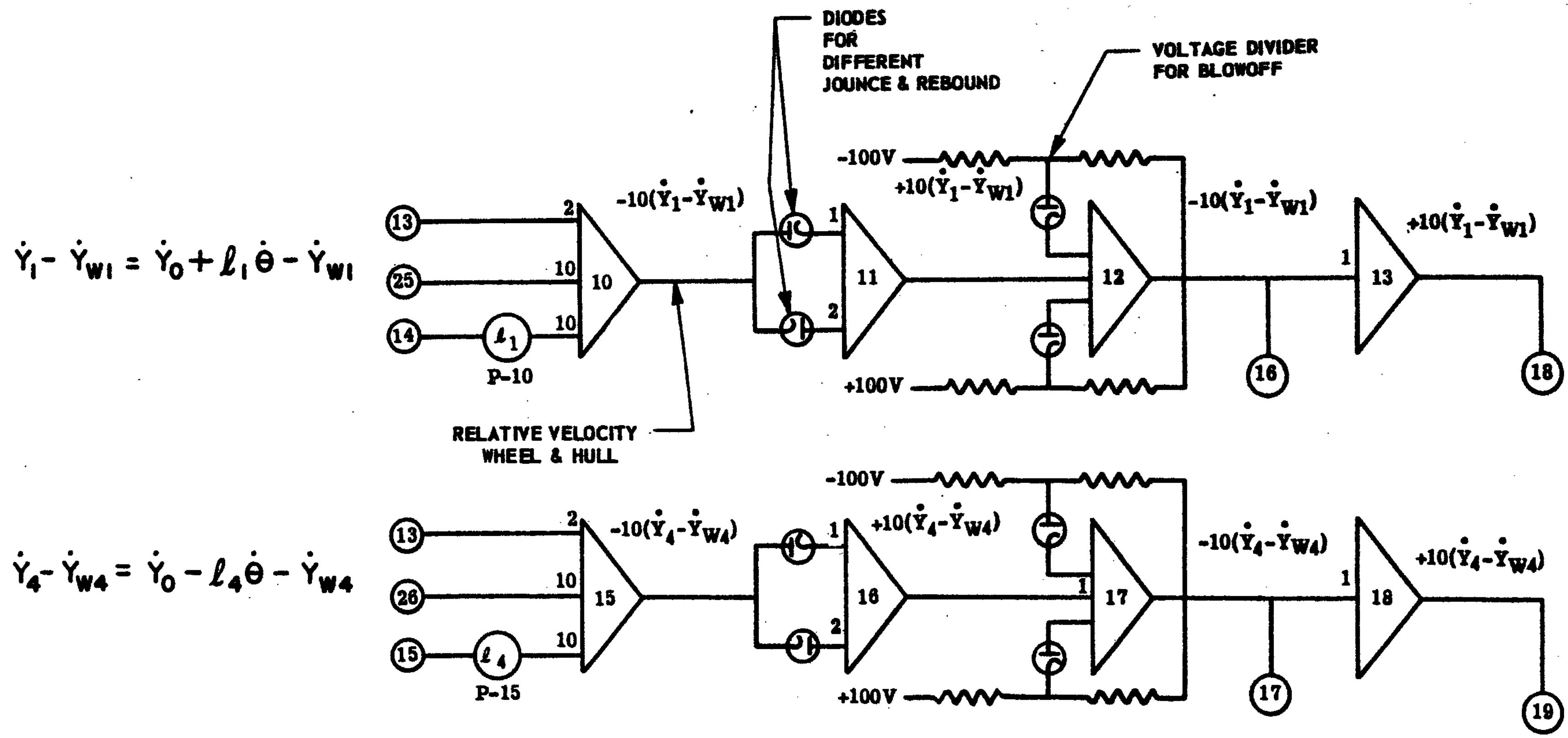


FIG. 9 SHOCK ABSORBER CIRCUITS

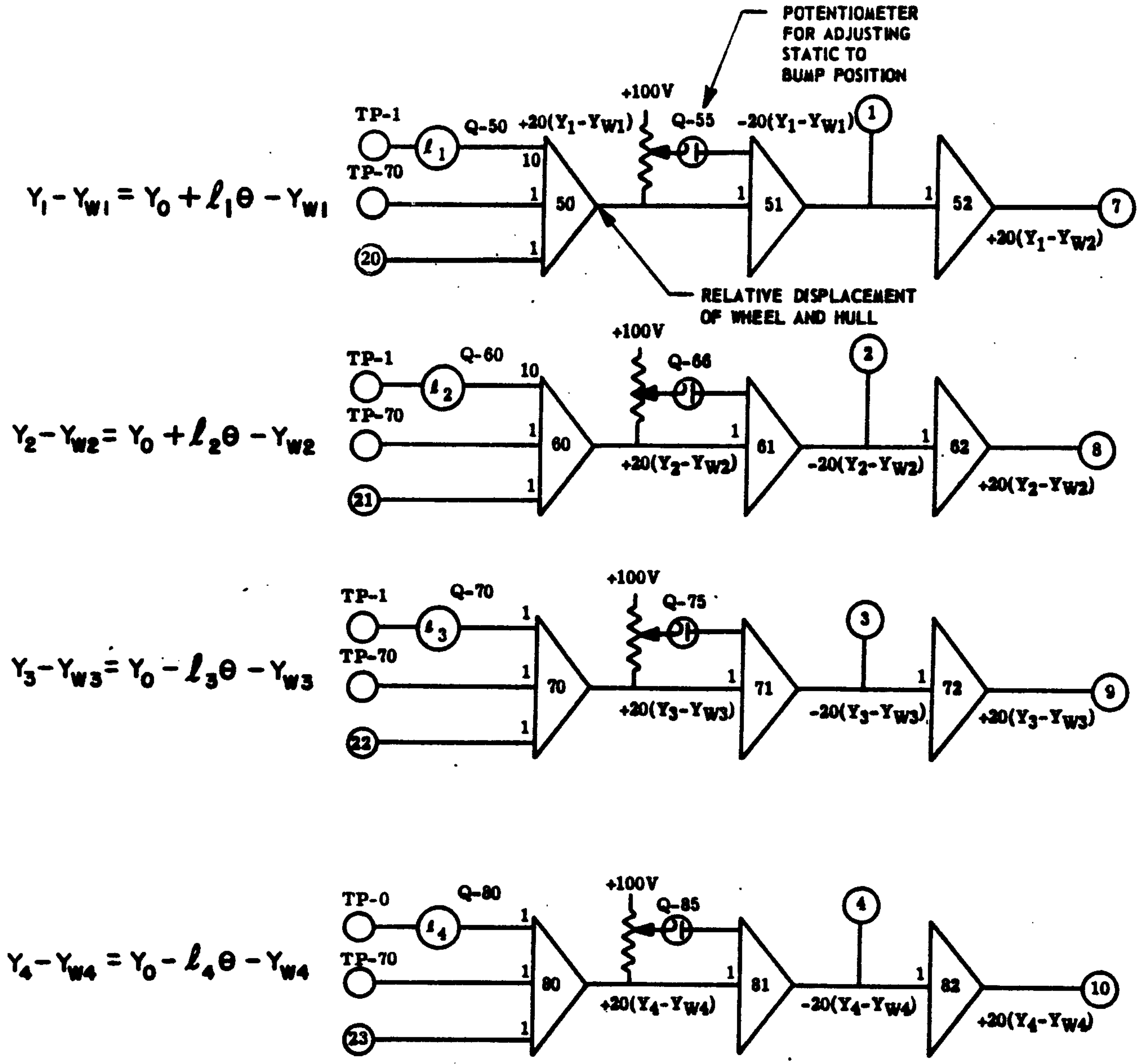


FIG. 10 SPRING CIRCUITS

$$\Sigma \frac{F_W}{M_W} = \ddot{Y}_W$$

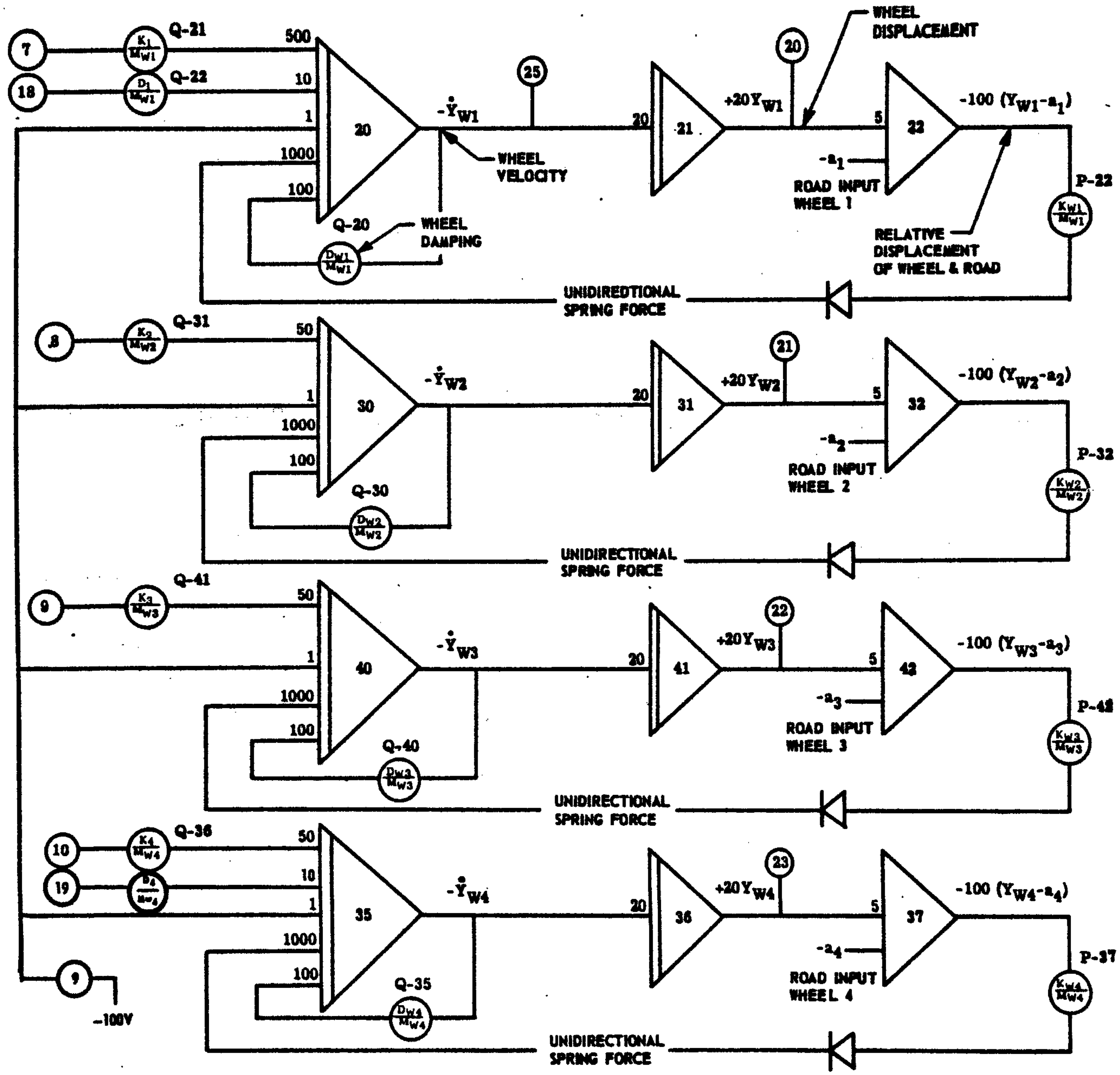


FIG. II WHEEL VELOCITY, AND DISPLACEMENT

$$\frac{\Sigma F_Y}{M_0} = \ddot{Y}_0$$

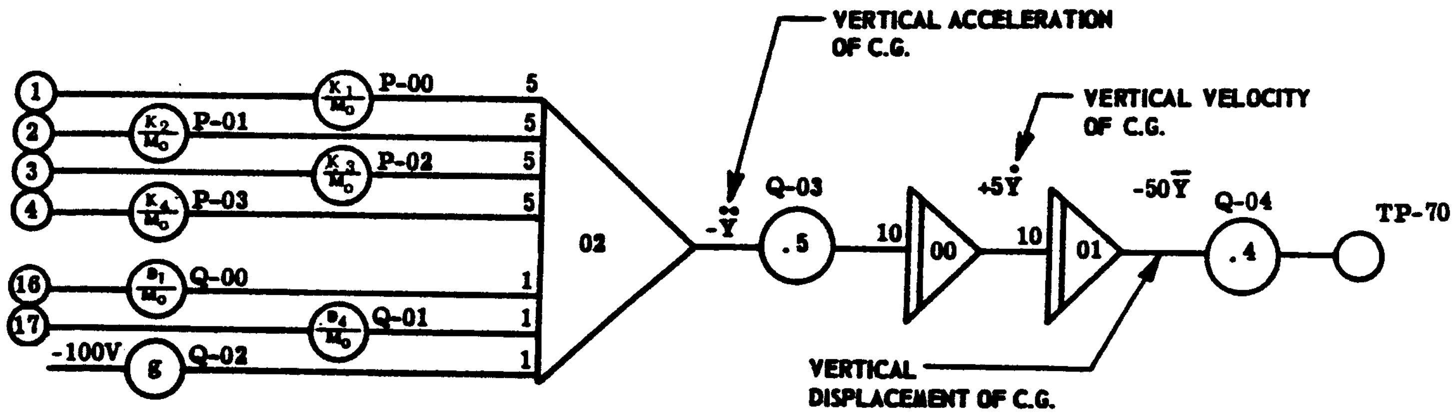


FIG. 12 C.G. VERTICAL ACCELERATION, VELOCITY, DISPLACEMENT

$$\frac{\Sigma T}{J_o} = \ddot{\theta}$$

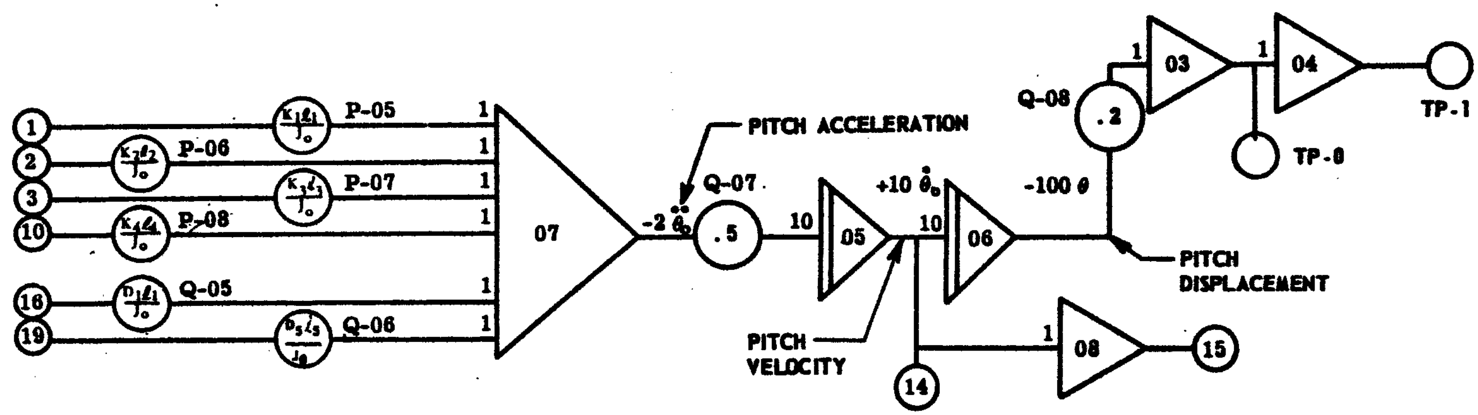


FIG. 13 C.G. PITCH ACCELERATION, VELOCITY, DISPLACEMENT

Simulation begins when the electronic version of a road and suspension system are brought together. Results are best analyzed using oscillographic or pen recorder output devices. Recorded paper tracings provide an excellent permanent record for lengthy detailed analysis. The oscillograph display system offers an opportunity to observe the simulation visually as an animated presentation. The dynamics of a complete vehicle or any component thereof may then be studied. This display system is used in conjunction with the analog computer. A cathode-ray-tube is used to convert the output voltages of the computer into a direct pictorial representation. Application of this system is shown in Figure 14. The series of photographs describe the motion of a tank that would be seen on the tube. The vehicle is shown negotiating at successive instances a 4" x 4" square obstacle.

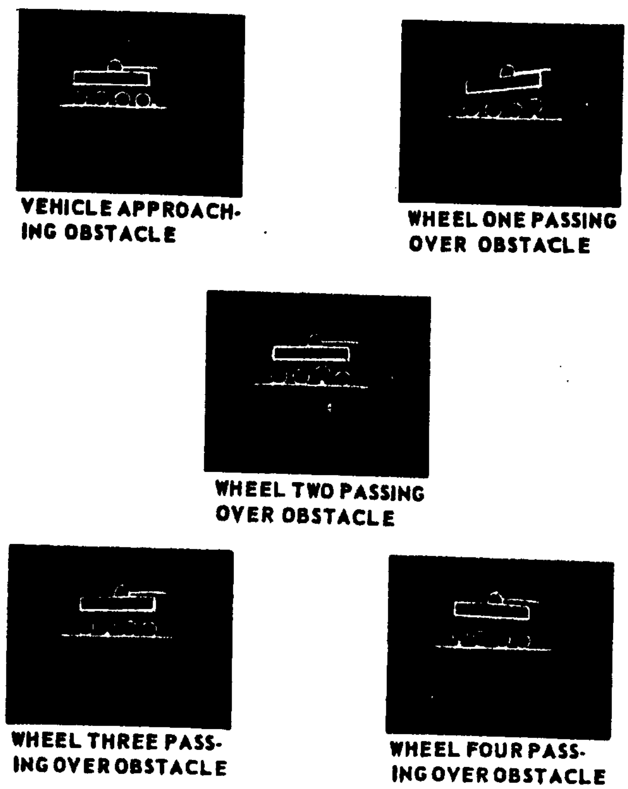


FIG. 14. VISUAL DISPLAY SYSTEM

This visual display provides a quick and easy method of conducting a preliminary analysis of new suspensions. It is also a good means of debugging a new simulation setup.

<u>**Simulator:**</u>

Results of computer simulations may also be studied with the aid of a motion simulator. The value of the Simulator, Figure 15, lies in its ability to physically reproduce realistic "ride motion" that may be predicted by a computer simulation. Thus, by combining computer studies of new concepts with a simulator analysis, design merits may be judged in the laboratory by engineers, designers, and administrative people before a design is considered for fabrication. Each individual may ride a new suspension in the Simulator and personally evaluate his area of interest firsthand.

The Simulator described below is a four degree of freedom machine capable of providing bounce, pitch, roll and yaw motions.

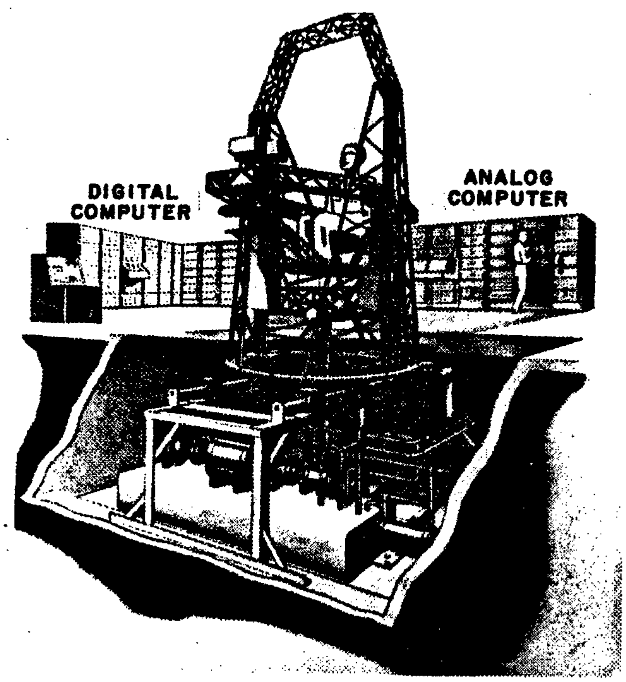


FIG. 15. SIMULATOR

The control of the machine is optional. The Simulator may be controlled from an instrument panel to produce either sine, square or triangular motions. Random motion inputs may be fed directly into the Simulator from an Analog computer simulation or by reproducing information previously recorded

on magnetic tape.

This machine is hydraulically driven and electronically controlled. Each of the four motions may be used individually or simultaneously.

Motion	Max. Tot. Travel	Max. Frequency	Acceleration
Bounce	3 ft	10 cps	2 g's
Roll	40 deg	10 cps	30 radians/sec^2
Pitch	40 deg	10 cps	30 radians/sec^2
Yaw	20 deg	3 cps	15 radians/sec^2

Perhaps the most significant claim that can be broadcast for the Simulator, at this time, is that it will make possible performance trials of designs prior to building of a design. In some instances it is the only economical approach, considering time and cost, particularly, where a new design is being investigated using many alternatives.

The creation of this Simulator provides the Army Tank-Automotive Command with a design tool that has been sought for some time. The need for an instrument of this kind has been in continuous demand for military suspension studies and other shock and vibration programs.

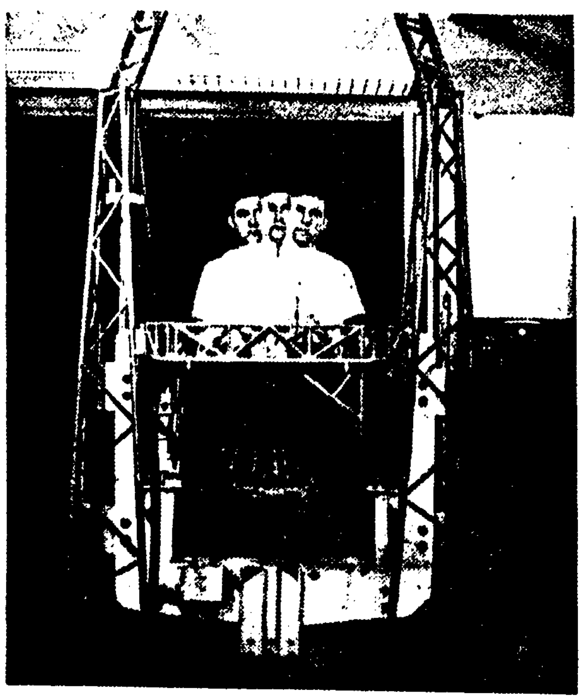


FIG. 16. DYNAMIC SIMULATION

The immediate response to the Simulator was generally favorable, but reserved. Comments usually indicated that the vibratory motions were good. However, it was repeatedly stated that the laboratory environment around the Simulator degraded the intended realism. The common complaint was that the "out of doors" atmosphere seemed to be missing.

To compensate for this a visual display was created providing a 180 degree field of view horizontally and 48 degrees in the vertical plane. A 35mm motion picture format was used to produce a "three screen" presentation. This method was selected, based upon successful tryouts of a unique projection system developed and tailored to the Simulator.

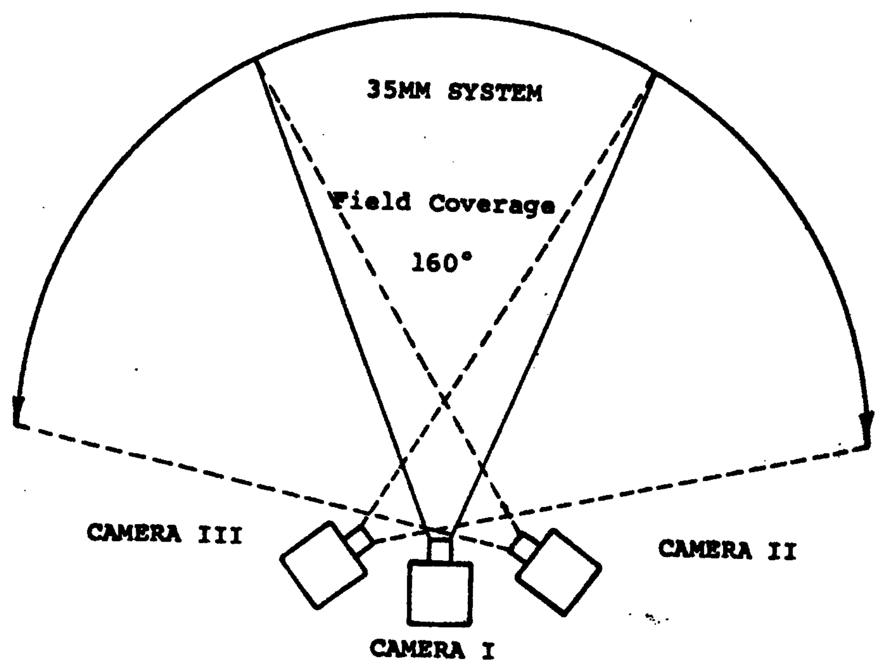


FIG. 17. CAMERA SYSTEM

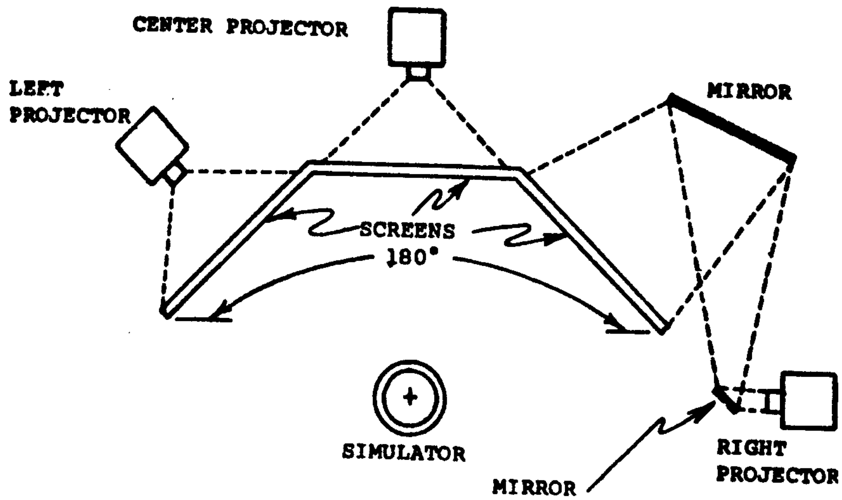


FIG. 18. PROJECTION SYSTEM

The activity scene is photographed by three cameras and backprojected to the subject in the Simulator by three synchronized-interlocked projectors. This system presents to the observer a scene that compares favorably to a view from within a moving vehicle.

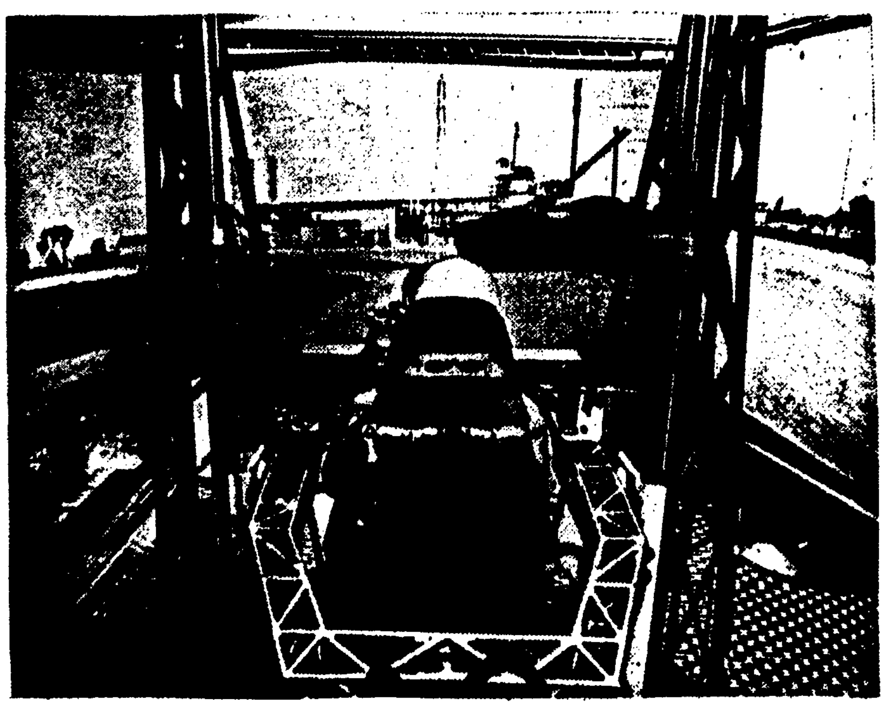


FIG. 19. DISPLAY SYSTEM

Vehicle Stability:

The Simulator was also used to simulate vehicle firing stability dynamics. The starting cue for this program was very forceably observed in vehicles like the Self-Propelled Artillery Weapon shown in Figure 20.

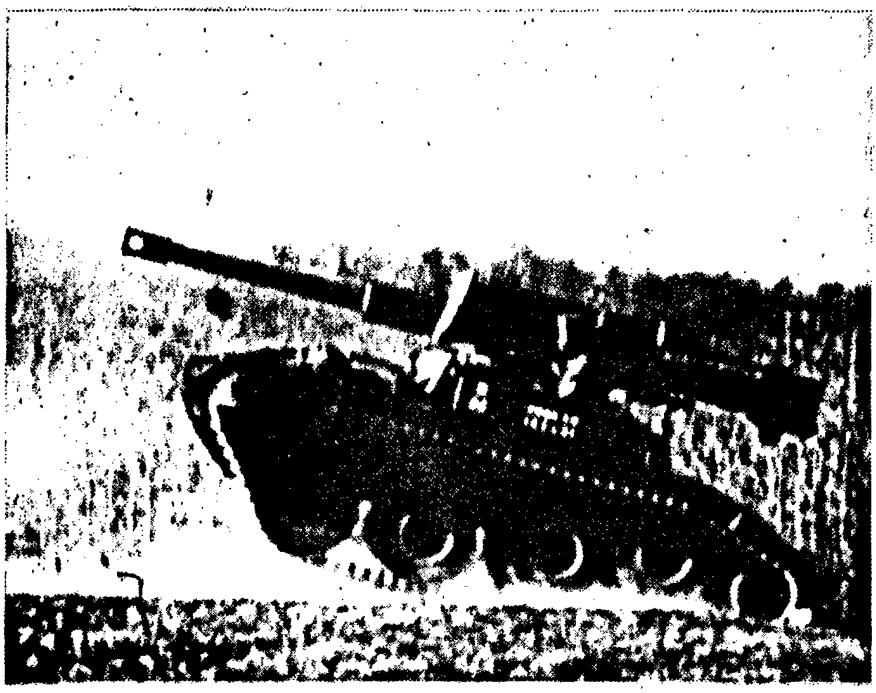


FIG. 20. SELF-PROPELLED ARTILLERY WEAPON

In keeping with this indicated trend of big guns on small chassis platforms it was necessary to establish with greater accuracy, the stability of contemplated designs.

The characteristic events describing fire stability were analyzed by charting the flow of events and establishing the equations of motion.

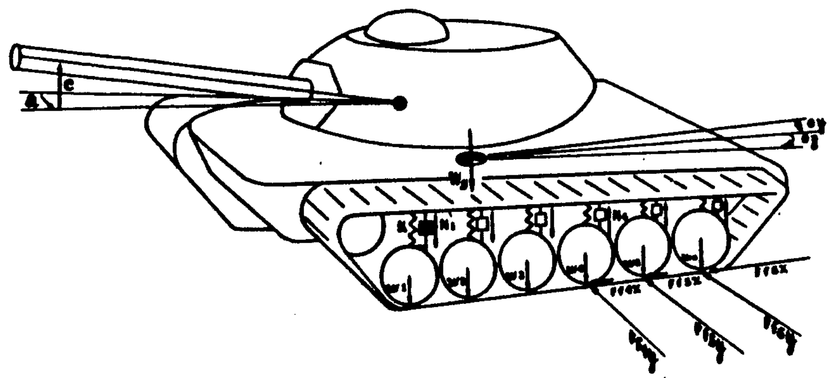


FIG. 21. VEHICLE SCHEMATIC

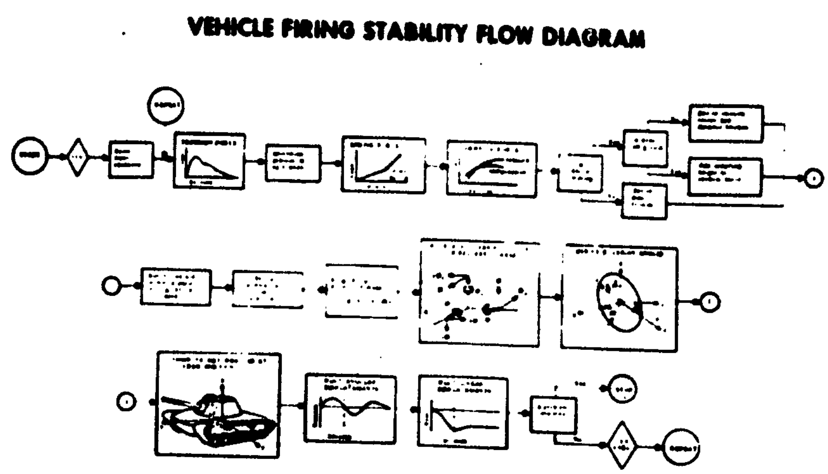


FIG. 22. FLOW DIAGRAM

Equations of Motion:

LONGITUDINAL TRANSLATION: GUN FORCE + GROUND FRICTIONAL FORCE = MASS X ACCELERATION $F_{gx} + F_{fx} = m\ (\ddot{x} - y\ \dot{\theta}_z + z\dot{\theta}y)$

LATERAL TRANSLATION: GUN FORCE + GROUND FRICTIONAL FORCE = MASS X ACCELERATION $F_{gy} + F_{fy} = m\ (\ddot{y} - \dot{z}\theta_x + \dot{x}\theta_z)$

BOUNCE: GUN FORCE - SPRUNG WEIGHT + SUSPENSION FORCES = MASS X ACCELERATION $F_{gz} - W_s c_z Z + \sum_{i=1}^{n} N_i = m\ (\ddot{z} - \dot{x}\dot{\theta}_y + \dot{y}\dot{\theta}_x)$

ROLL: GUN MOMENT + GROUND FRICTIONAL MOMENT + SUSPENSIONAL MOMENT = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{z}F_{gy} + \bar{y}F_{gz} + (z + z_0)\ F_{fy}c_yY + \sum_{i=1}^{n} N_iY_i = \ddot{\theta}_xI_x - \ddot{\theta}_zI_{xz} +$

$$(I_z - I_y)\ \theta_y\dot{\theta}_z - I_{xz}\dot{\theta}_x\dot{\theta}_y$$

PITCH: GUN MOMENT - GROUND FRICTIONAL MOMENT - SUSPENSIONAL MOMENT = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{x}F_{gz} + \bar{z}F_{gx} - (z + z_0)\ F_{fx}c_xX - \sum_{i=1}^{n} N_ix_i = \ddot{\theta}_yI_y + \dot{\theta}_x\dot{\theta}_x\ (I_x - I_z)$

$$+ (\dot{\theta}_x^2 - \dot{\theta}_z^2)\ I_{xz}$$

YAW: GUN MOMENT + GROUND FRICTIONAL MOMENTS = ANGULAR ACCELERATION X MOMENT OF INERTIA $-\bar{y}F_{gx} + \bar{x}F_{gy} + \sum_{i=1}^{n}$

$$x_iF_{fiy}\ c_yY - \sum_{i=1}^{n} y_iF_{fix}c_xX = \ddot{\theta}_zI_z - \ddot{\theta}_xI_{xz} +$$

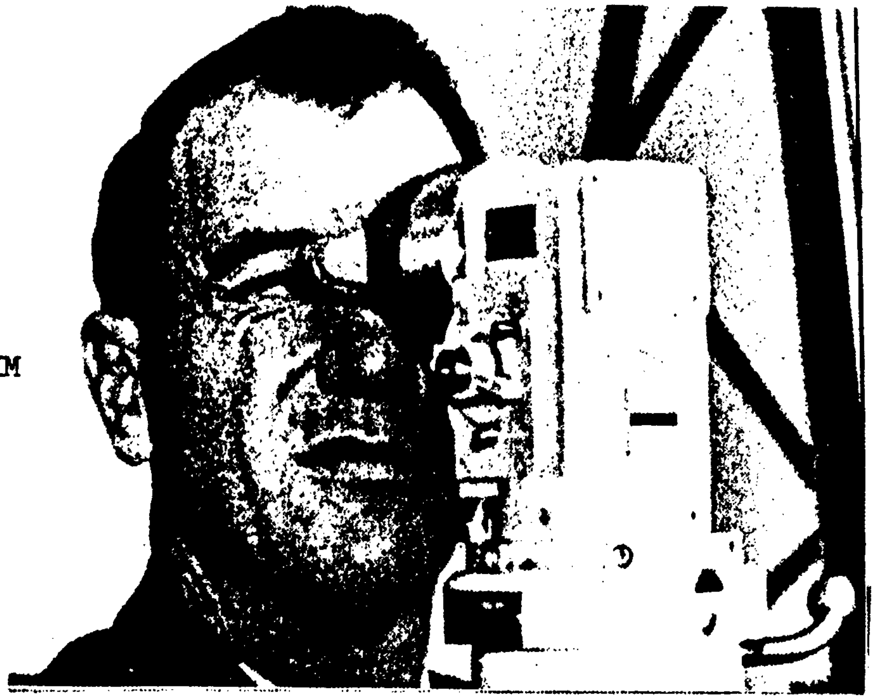
$$(I_y - I_x)\ \dot{\theta}_x\dot{\theta}_y + I_{xz}\dot{\theta}_y\dot{\theta}_z$$

The derived statements were for weapon systems free to move in three degrees of angular freedom - roll, pitch and yaw; and three degrees of translational freedom - fore and aft movement, bounce, and lateral slip. The equations define vehicle motion as effected by interrelated factors of gun firing force, gravity, terrain influence, and the resisting forces of the suspension.

The dynamics of firing stability are calculated on a digital computer. When this program is used in conjunction with the Simulator, the results are stored in computer memory. The physical arrangement of the Simulator permits the occupant

to fire any weapon by merely pulling the usual trigger. The command to fire is completely controlled by the man in the seat.

AIM



FIRE

FIG. 23. FIRING SIMULATION

The inputs to this problem consist of various vehicle measurements, weights, moments of inertia, gun recoil force, type of soil, and type of suspension (active or locked-out). The output consists of detailed information in tabular or graph form showing angular and translational disturbances and their respective displacements, velocities, and accelerations with respect to time. This information describing gun firing force impact on the vehicle and resultant vibrations is available for the C.G. of the vehicle with reference to earth-fixed axes and for any other point on or within the vehicle, such as gun muzzle, engine mounts and crew stations with reference to the vehicle axes.

Summary:

The combination of computers and simulation techniques at the Army Tank-Automotive Command has provided an effective and versatile designer's tool. Informative preliminary studies have been conducted of new suspension systems and stability characteristics, without the use of hardware units of the design. Probes of unique approaches have quickly established design direction and payoff areas.

SUSPENSION NOMENCLATURE

$\ddot{Y}_o$ = Vertical acceleration of the center of gravity.

$\dot{Y}_o$ = Vertical velocity of C.G.

Y_o = Vertical displacement of C.G.

$\ddot{\theta}_o$ = Pitch acceleration about C.G.

$\dot{\theta}_o$ = Pitch velocity about C.G.

θ_o = Pitch displacement at C.G.

$(Y_1 - Y_{w1})$ = Relative displacement between hull and wheel at wheel 1.

$(\dot{Y}_1 - \dot{Y}_{w1})$ = Relative velocity of the hull and wheel at wheel 1.

$(Y_{w1} - a_1)$ = Relative displacement between wheel and input bump at wheel 1.

J_o = Pitch Moment of Inertia.

M_o = Sprung mass.

M_w = Wheel mass.

ℓ = Distance from wheel centerline to C.G.

K_{1-4} = Suspension spring constant.

D_{1-4} = Shock absorber damping constant.

K_w = Spring constant of road wheel rubber.

D_w = Damping constant of road wheel rubber.

a_{1-4} = Road inputs to wheel No. 1-4.

Y_{1-4} = Chassis displacement.

g = Acceleration of gravity.

VEHICLE STABILITY NOMENCLATURE

F_g = Gun force.

F_x = Frictional force.

W_s = Sprung weight.

M = Sprung mass.

$\bar{x}, \bar{y}, \bar{z}$ = Position of trunnion centerline.

z_o = Static height of C.G.

n = Number of wheels.

$\ddot{y}, \ddot{x}, \ddot{z}$ = Translational acceleration.

$\ddot{\theta}_x$ $\ddot{\theta}_y$ $\ddot{\theta}_z$ = Angular acceleration.

$\dot{y}$ $\dot{x}$ $\dot{z}$ = Translational velocity.

$\dot{\theta}_x$ $\dot{\theta}_y$ $\dot{\theta}_z$ = Angular velocity.

I = Moment of Inertia.

K = Suspension spring constant.

D = Shock absorber damping.

a = Gun azimuth.

e = Gun elevation.

ACKNOWLEDGMENTS

Acknowledgment is extended to the past and present members of the Dynamic Simulations Laboratory at the Army Tank-Automotive Command for their individual contributions which led to the completed development described in this paper, and to Mr. O. K. Till, F. Anderson and R. Armstrong for development of the experimental equipment and instrumentation. I am also indebted to Mr. J. Stanton and J. Lynch, Jr. for their art and photographic work in preparing the illustrations.

REFERENCES

1. S. Heal, "Suspension Analysis", Army Tank-Automotive Command, Research Report No. RR-38.

2. M. Archambault, S. Heal, S. Staich, "Generation of Road Profiles for Vehicle Ride Simulations", Army Tank-Automotive Command, Research Report No. RR-44.

3. C. Fischer, "Mathematical Model and Digital Computer Program for Vehicle Firing Stability Analysis", Army Tank-Automotive Command, Research Report No. RR-27.

4. I. J. Sattinger, S. Sternick, "An Instrumentation System for the Measurement of Terrain Profile", University of Michigan Report No. 2948-36-F, December 1961.

5. "Operating and Maintenance Instructions for Simulation Display System", University of Michigan Report No. 2023-518-M, October 1955.

6. I. J. Sattinger, D. F. Smith, "Computer Simulation of Vehicle Motion in Three Dimensions", University of Michigan, May 1960.

7. R. Dean - Averns, "Automobile Chassis Design".

8. V. Kowachek, "Firing Stability", Technical Notes.

SIZE EFFECTS IN THE MEASUREMENT OF SOIL STRENGTH PARAMETERS

Ben Hanamoto
Land Locomotion Laboratory
Army Tank and Automotive Center

and

Emil H. Jebe
Institute of Science and Technology
The University of Michigan

INTRODUCTION. The main concern of the Land Locomotion Laboratory, ATAC, is the relationship of a vehicle to the terrain over which it travels. Once an insight into this relationship has been attained, the problems encountered from the initial design of a vehicle to its ultimate use in the field can be more easily grasped and rationally solved. The solution of engineering problems depends upon the selection of the relevant variables and a description of the functional relationship among these variables. The selection of the vehicle variables are within broad limits at the disposal of the designer, but for the terrain or soil, the selection of suitable variables becomes much more complicated. Soil is probably one of the most complex of all engineering materials [1]. Researchers in soil mechanics have added much to the knowledge of the mechanical and physical characteristics of soils, but as yet, no fully satisfactory general theory is available.

Land locomotion is an engineering application of soil mechanics to off-road vehicular operation. Its objective is to determine the relationships of vehicles, or more precisely, of the wheel and track, to the strength properties of the soil. In land locomotion research, one of the important questions is the nature of the pressure-sinkage relationship. The Bekker equation

$$p = \left(\frac{k_c}{b} + k_\phi \right) z^n \qquad (1)$$

represents a family of curves with three unknown constants, k_c, k_ϕ, and n, and two variables, p the pressure under a loaded area, and z the sinkage* [2].

* b is a known constant, the plate width. p is per unit area (i.e., square inch).

A set of these constants which will approximately describe pressure-sinkage observations can be obtained [3]. To determine these constants, pressure-sinkage experiments were performed in the laboratory with footings of various sizes and shapes. The constants obtained were used to predict the sinkages for other loaded areas. It was found that the predictions were adequate for tracked vehicles, where the relative shapes of the loaded areas were similar. Rectangular test footings with a length/width ratio greater than 5 were used for determining the soil parameters used in the prediction equations. The basic equation includes only the width term b. For cohesionless soils, the ultimate bearing strength is dependent on the width only for long loaded areas [4]. To determine what the minimum length had to be before a footing was not considered long, laboratory tests were conducted with footings of varying ℓ/b ratios. Acceptable results for the pressure-sinkage relationship were obtained when the ℓ/b ratio was greater than 5. Consequently, all pressure-sinkage measurements were taken with footings of at least an ℓ/b ratio of 5 or greater.

The ultimate equations in which these soil strength parameters are to be used apply to the general case of predicting vehicle performance for both tracked and wheeled vehicles. As noted above, the predictions of tracked vehicle sinkage and motion resistance have been generally satisfactory. For wheeled vehicles, however, improvements are needed. One of the differences to be noted between a tracked and wheeled vehicle is the shape of the loaded area. In most cases, tracked vehicles have a contact area of relatively long length as compared to width. Such a length-width ratio is not the usual situation for wheeled vehicles at moderate sinkages. Most tires will have a contact area where the ℓ/b ratio is close to 1 or 2.* Therefore, it was thought that the shape of the loaded area when ℓ/b was less than 5 might have effects on the pressure-sinkage relationship. A clearer understanding of the pressure-sinkage relationship in this region would provide us with an improved soil-vehicle model with broader and more useful applications. Consequently, a test program was undertaken by the Land Locomotion Laboratory to investigate further the pressure-sinkage relation.

---

*For example, the Army 6 x 6 5-ton truck normally carries an 11.00 x 20 tire. At one inch sinkage, ℓ/b - 1 and at eight inch sinkage ℓ/b = 2.33.

DESCRIPTION OF TEST PROGRAM. The test program to study the effect of plate size on the vertical soil strength or sinkage parameters was divided into two parts. The first part comprised a study of the reliability or repeatability of the test results for the equipment and mixing techniques that were to be used in the tests. The second part covered the measurement of the load-sinkage curves while using different sized plates.

In carrying out the experiments a laboratory model bevameter is used to drive a constant speed hydraulic piston arrangement which pushes a plate into the soil in a bin. The depth of sinkage and the force on a load cell are plotted electrically by an X-Y plotting device. The curve is traced on linear graph paper. From this graph values for p and z may be read off and plotted on log-log paper. The slope of the least square fitted line on the double-log plot gives an estimate of the parameter n. The constant term in such a fitted equation is the logarithm of

$$\frac{k_c}{b} + k_\phi$$

where b is the width of the plate used. Thus it is seen that k_c and k_ϕ cannot be estimated by any straight forward statistical technique. Use of two different b values, however, will permit setting up two simultaneous equations in k_c and k_ϕ.

PART I. We wished to determine the maximum number of penetrations that could be made with one preparation of the soil bin. For this purpose we set up a uniformity trial using only a 2" x 10" plate. Orientation and location of the penetrations was arranged as indicated in Figure 1. The three orientations shown were carried out in a randomized block design with six replicates.

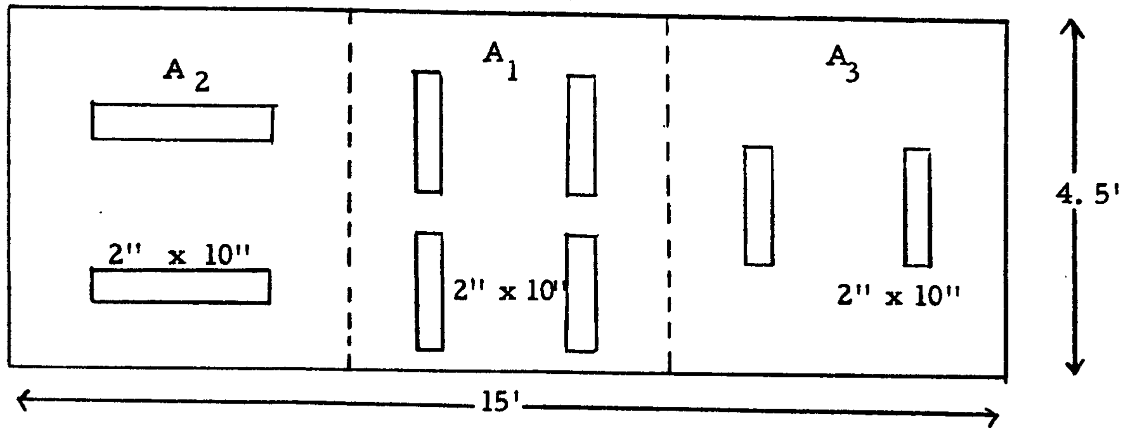


FIGURE I: SAMPLE BLOCK LAYOUT FOR UNIFORMITY TRIALS

RESULTS FOR PART I: Mean pressures were computed for the orientations at 1", 2" and 3" depths of sinkage. Mean differences for the orientations at a given depth were found to be homogeneous (i. e., not significant). At the 2" depth of sinkage, the coefficient of variation was about 6 percent, quite a satisfactory value. Examination of the variability within the orientations showed that the variation within the arrangement A_1 was significantly greater than that for the other arrangements.

CONCLUSIONS FOR PART I: The experimental procedure would yield results of adequate reliability. The orientation A_1 appeared undesirable for taking pressure-sinkage measurements. Therefore, we decided to make the spacings between determinations as similar to the A_2 arrangement as practicable. Since a rectangular plate of size 3" x 10" was to be used in Part 2, it appeared necessary to "beef up the apparatus" to handle the greater pressures required to sink such a large plate.

PART 2. The second phase of the test program comprised measurement of the load-sinkage relation in dry sand with plates of varying sizes. Analysis of test data would provide estimates of the three parameters in the Bekker equation and their variations for the sizes and shapes of plates used. The results should show the dependence, if any, of the parameters on the size and shape of plate.

Many problems arose in the consideration of the second part of the test program. One point was that all plates should be tried in one mix of the bin. Thus, a complete block design would be preferred.

A second point is that the sinkage equation is a stress-strain relationship. It may also be described as a functional relation [5]. The problems of estimation which arise for the functional relation have been resolved by Dr. Joseph Berkson by introducing the "control variable" concept [6]. It appears that z may be taken as the control variable in our problem. This approach is contrary to the usual dependent-independent variable point of view, but Berkson has shown that the method is unbiased for estimating the functional relation if the errors in p are unbiased.

Thirdly, it was clear that a statistical analysis of the estimation procedure for the Bekker equation was needed. With transformation to logarithms of p and z it is assumed that the standard linear regression assumptions are valid in the transform [7]. Thus, estimation of the parameter n is quite straightforward. When n has been obtained, the procedure takes $z = 1$, hence, $\log z = 0$, and predicts a value of log p, say P_o. Now, anti-log $P_o = p^* = k_\phi + k_c/b$. By taking two values of the plate width, b_1 and b_2, and corresponding p_1^* and p_2^* values, the estimation equations for k_c and k_ϕ become

$$(2) \qquad k_c = \frac{(p_1^* - p_2^*)b_1b_2}{b_2 - b_1}$$

and

$$(3) \qquad k_\phi = \frac{b_2p_2^* - b_1p_1^*}{b_2 - b_1}$$

These results pointed out two things:

1. Widely spaced b values to permit use of large $b_2 - b_1$ would reduce the variance of the estimates of k_c and k_ϕ,

2. A formula for the variance of p^* is needed.

The variance formula for p^* presents some difficulty because p^* is a nonlinear function of P_o. We recall, however, that the choice of an appropriate experimental design will give us direct estimates of the experimental error for our estimated k_c and k_ϕ values so that we can bypass the variance formula problem.

DESIGN OF THE PART 2 EXPERIMENTS. From the statistical analysis it was clear that the largest possible difference in plate widths should be used to estimate the parameters k_c and k_ϕ. This consideration along with the desire for a complete block experiment already mentioned resulted in a revision of the choice of dimensions for the rectangular plates to be used in the experiment. The plate sizes actually selected are given in Table 1.

TABLE 1
LIST OF PLATE DIMENSIONS FOR PART 2 EXPERIMENTS

(in inches)

Rectangles:			
	1 x 4	2 x 4	3 x 4
	1 x 6	2 x 6	3 x 6
	1 x 8	2 x 8	3 x 8
	1 x 10	2 x 10	3 x 10
Circles:			
Diameters:	2 and 4.		

The entire set of plates thus provided 14 treatments to be set up in a completely randomized block experiment where one block equals a mix of the soil bin. Each plate was to be used in a randomly selected plot of size about 2' x 2' to make a single measurement of the pressure-sinkage curve. It was further decided to complete six replicates which would yield 24 independent pairs of estimates of k_c and k_ϕ, but would, of course, yield 84 estimates of the parameter n, one estimate being obtained from each pressure-sinkage curve.

ANALYSIS OF RESULTS FOR PART 2. Analysis of the estimated values for n is straightforward and results are readily interpreted. For the k_c and k_ϕ values some difficulties arise. Plotting the means is a most useful device for aid in understanding the effects indicated by the analysis.

Figures 2 and 3 show the width and length effects on n without considering the interaction. In order to present the interaction effect, we show the usual type of plot for displaying an interaction. Parallel lines for the various lengths of plate would be indicative of no interaction.

Thus, the lack of parallelism exhibited in Figure 4 is indicative of the nature and source of the interaction. The major pattern, however, is still that shown in Figures 2 and 3. There appears to be a decrease in the n values with an increase in area of plate whether the area increase is due to change in length or change in width. The width effect is much smaller in going from 2" to 3", but the length effect is about the same at all widths.

In considering the analysis of the estimated k_c and k_ϕ values it will be useful to recall equations (2) and (3) which show how the estimates are obtained. There are some obvious points to be noted from these equations, but we shall defer them until later. Each experimental unit or plot in the soil bin yields a p versus z relation as drawn by the x, y plotter while a single plate is sunk into the soil. Then results from two different plate widths have to be combined to obtain a single estimate of k_c or k_ϕ. We may combine 1" and 2" widths or widths of 2" and 3" or 1" and 3". As shown above, the latter is the best choice, but we see that a single replicate or "set" will only give us four such independent estimates.

By keeping the width difference constant and varying the length we can pair within one set these four pairs of plates:

1 x 4	and	3 x 4	1 x 8	and	3 x 8
1 x 6	and	3 x 6	1 x 10	and	3 x 10

to give us four pairs of estimates for k_c and k_ϕ. Utilizing the six "sets" gave us 24 values for analysis.* Estimates obtained from these pairings may be analyzed for the length effect alone. But we would also like to study the width of plate effect on these two parameters and look for interaction, if any, as we did for the parameter n.

First, we consider the estimates of k_c. The mean values obtained for the pairs just listed were:

*A set consists of six replicates carried out at the same depth. Eighteen replicates were actually completed at each of two depths. Aggregation over six replicates forms a set. Thus, there are a total of six sets.

Figure 2

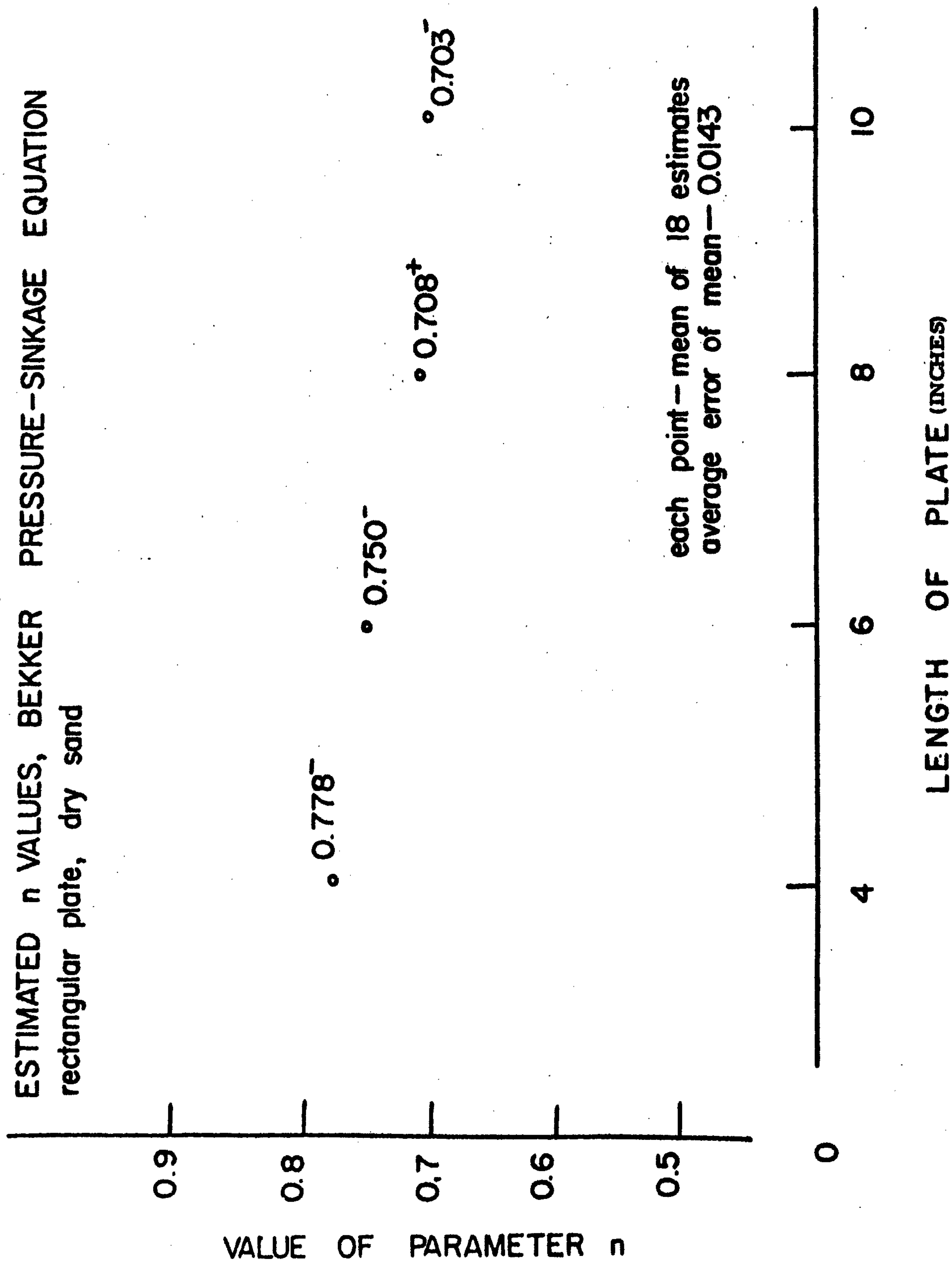

ESTIMATED n VALUES, BEKKER PRESSURE-SINKAGE EQUATION
rectangular plate, dry sand
each point—mean of 18 estimates
average error of mean—0.0143
0.778⁻
0.750⁻
0.708⁺
0.703⁻
0.9
0.8
0.7
0.6
0.5
VALUE OF PARAMETER n
0
4
6
8
10
LENGTH OF PLATE (INCHES)

Figure 3

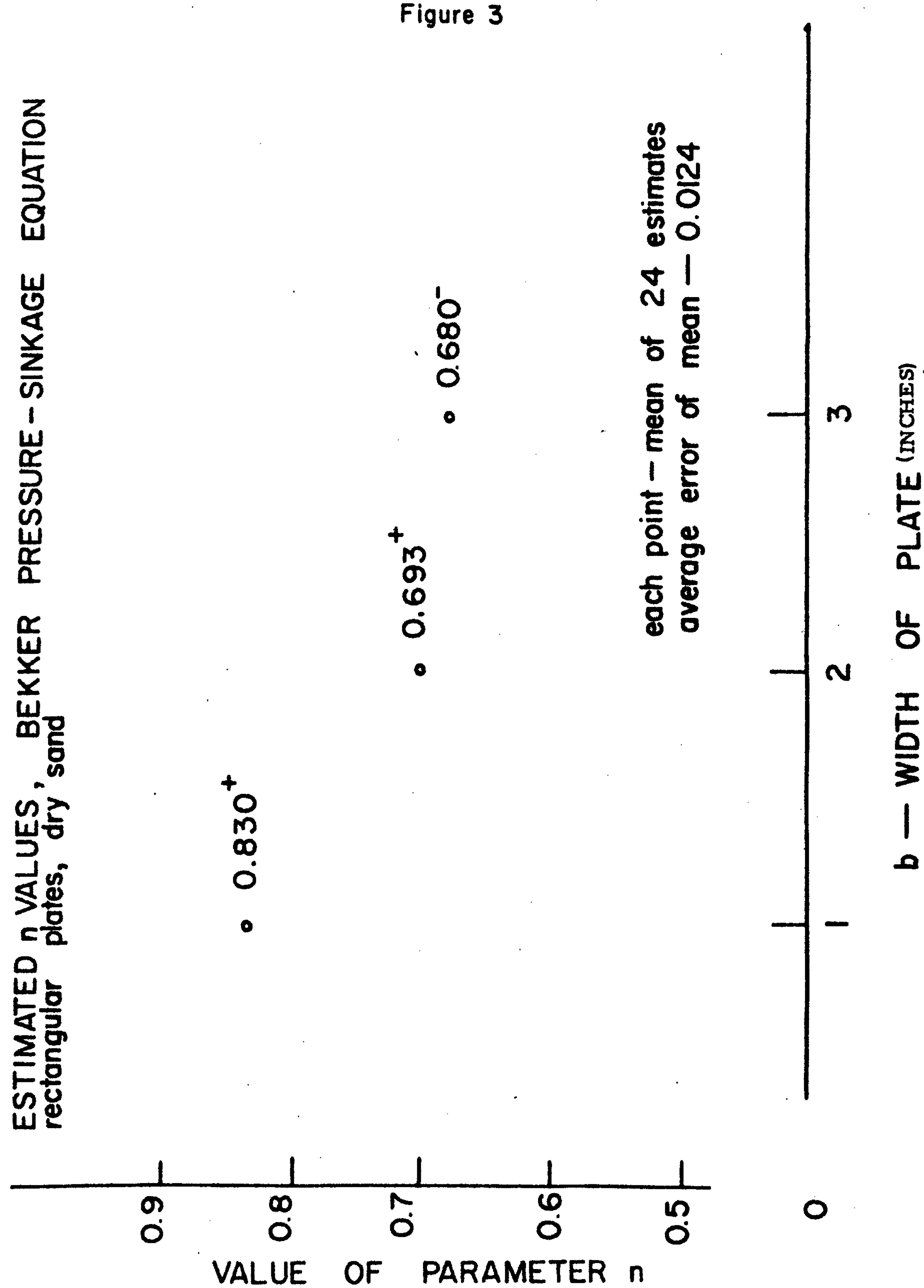
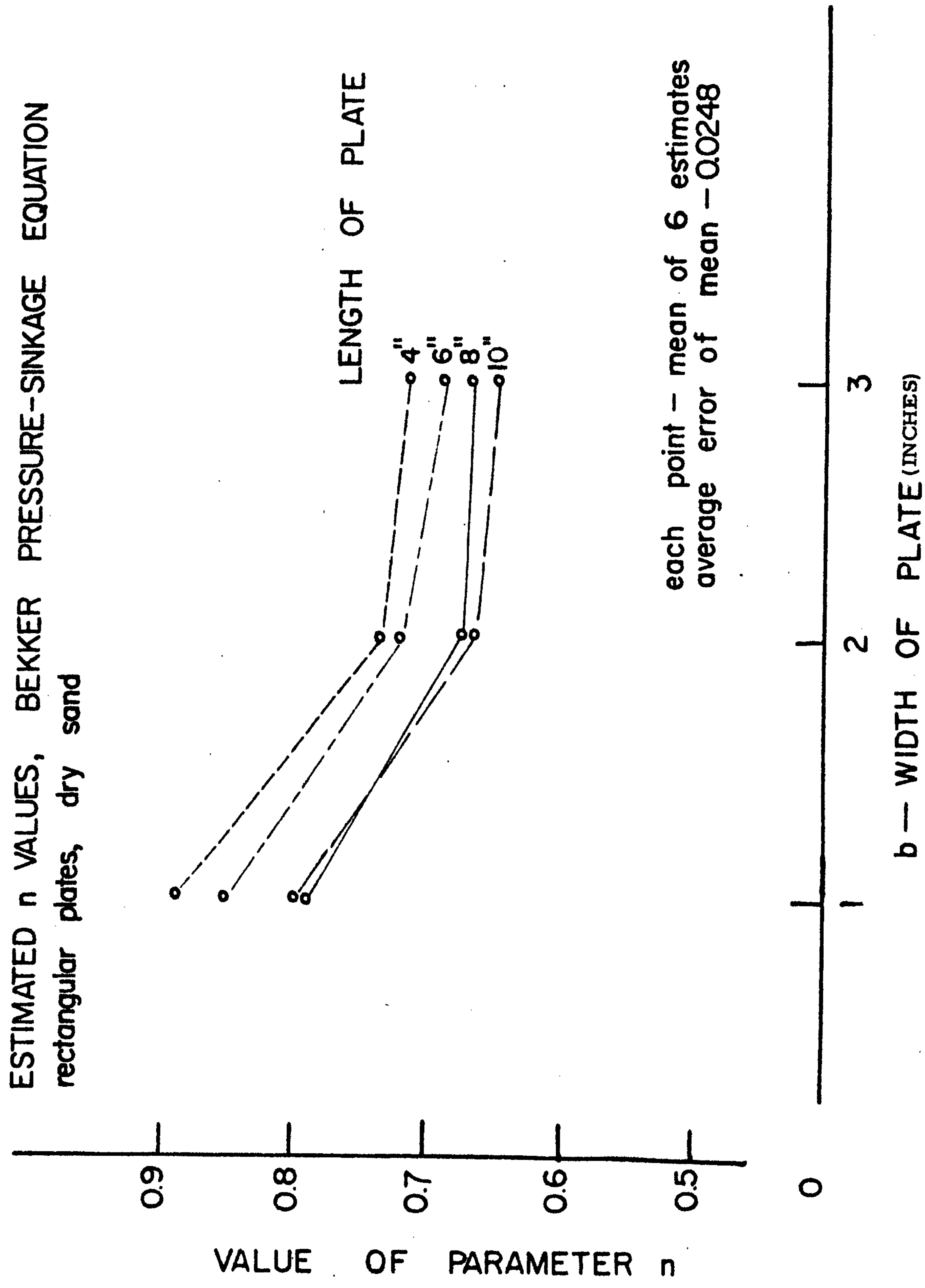

ESTIMATED n VALUES, BEKKER PRESSURE - SINKAGE EQUATION
rectangular plates, dry sand
0.830+
0.693+
0.680-
each point - mean of 24 estimates
average error of mean — 0.0124
0.9
0.8
0.7
0.6
0.5
VALUE OF PARAMETER n
0
1
2
3
b — WIDTH OF PLATE (INCHES)

Figure 4


ESTIMATED n VALUES, BEKKER PRESSURE-SINKAGE EQUATION
rectangular plates, dry sand
LENGTH OF PLATE
4"
6"
8"
10"
each point - mean of 6 estimates
average error of mean - 0.0248
0.9
0.8
0.7
0.6
0.5
VALUE OF PARAMETER n
0
1
2
3
b — WIDTH OF PLATE (INCHES)

Pairs:

1 x 4	1 x 6	1 x 8	1 x 10
and	and	and	and
3 x 4	3 x 6	3 x 8	3 x 10

Mean Value of k_c

-3.305	-4.030	-4.000	-4.166

Relevant mean squares abstracted from the analysis of variance [8] for comparing these means are:

Source	d. f.	M. S.
Length of Plate	3	0.8977
Length of Plate by Depth	3	1.0526
Length of Plate by Sets within Depth	12	0.2262

Since the $P(F_{3,12} \geq 3.89) = 0.05$, we may conclude that there is a length effect on k_c and that the length effect is not uniform at the two depths studied (i. e., $F = 3.97 = 0.89770.2262$).

To extend our analyses other plate pairings were studied. Finally, it seemed in our judgment that the following pairs were most useful for securing information on the width effect plus a little more length information:

Pairs:

1 x 4	1 x 6	2 x 4	2 x 6
and	and	and	and
2 x 8	2 x 10	3 x 8	3 x 10

Mean Value of k_c

-2.463	-2.495	-5.169	-5.063

Note that the pairs selected show a constant length difference of 4 inches. Furthermore, the length differences are balanced for comparing 1 and 2 inch widths with 2 and 3 inch widths for the study of the plate width effect in estimating k_c. It appears that the latter effect is large. Relevant mean squares from the analysis of variance are as follows:

Source	d. f.	M. S.
Width Pairs	1	41.717
Width Pairs by Depth	1	7.183
Length Differences within Width Pairs		
Length in 1 and 2	1	0.003
Length in 2 and 3	1	0.034
Length Difference by Depth		
in 1 and 2	1	0.006
in 2 and 3	1	5.914
Experimental Error	12	1.529

For comparing the width pairs we obtain an F ratio = 41.717/1.529 > 27. We conclude that the parameter k_c depends on the choice of plate widths. In estimating k_c, pairing of 1 and 2 inch widths gives a larger value (algebraically) than pairing 2 and 3 inch widths. Looking at it another way the data show that changing the length/width ratio for the plates from

the range 4 to 6 to a range of 2 to 3 materially alters the k_c value obtained.

Similar analyses for the estimates of the parameter k_ϕ were carried out as just described for k_c. We note that the k_ϕ estimates were positive, mostly in the range +7 to +10. Our conclusions about the effect of plate width on the parameter k_ϕ were the same as reached in regard to k_c, but no length effect was detected.

With these results available we may return to the further consideration of the method used to obtain the estimates. First, we observe that if there were no width of plate effect in the p versus z relation, then k_ϕ would be equal to, say, p^*. In the notation used above, no width of plate effect would mean $P_{01} = P_{03}$, if one inch and three inch plate widths were used, and, hence, $p_1^* = p_3^* =$ a common value, p^*. Observed p_1^* and p_3^* values would differ, of course, due to experimental variation, but average values would be equal. Further, when $k_\phi = p^*$, then the solution for k_c is zero. The parameter k_c was once regarded as a measure of cohesiveness of the soil and, hence, might be zero for sand [3]. The present experiments, in which mason's sand was used certainly do not agree with this point of view about the parameter k_c.

Recognizing then that the present model for the p versus z relation does admit a width of plate effect our analyses so far have indicated the magnitude of this effect. In addition, we have obtained some indication of a length effect on k_c although none appeared for k_ϕ.

Second, with respect to the method of obtaining the estimates of k_c and k_ϕ, we point out that the estimates obtained are correlated. This correlation cannot be avoided because the values are based on the solution of two simultaneous equations. In fact, the estimates also would have been correlated if we had been able to obtain them by a direct least squares procedure. In the least squares case, however, it is usually easy to write down the covariance of the estimates and, hence, to find the correlation if desired. For the two simultaneous equations in k_c and k_ϕ, matrix methods can be applied to find the covariance matrix for k_c and k_ϕ. The result obtained for the covariance of k_c and k_ϕ is

$$(4) \qquad \mathrm{cov}(k_c, k_\phi) = (b_2 - b_1)^{-2} (b_1^2 b_2 V(p_1^*) + b_1 b_2^2 V(p_2^*))$$

where b_2 is the width of the larger plate and V( ) denotes variance of the enclosed quantity.

It is interesting to examine the consequences for different choices of plate widths in estimating k_c and k_ϕ. As a first approximation it may be adequate to assume $V(p_1^*) \cong V(p_2^*)$. If this assumption is made, the following correlation values are obtained:

Plate widths paired: (widths in inches)		
1 and 3	1 and 2	2 and 3
Value of correlation of k_c and k_ϕ		
-0.8944	-0.9487	-0.9805

Observed values of the correlation between k_c and k_ϕ will, of course, differ from these theoretical values because of experimental variation and lack of equivalence of $V(p_1^*)$ and $V(p_2^*)$. It is clear, however, that the above results further support the use of maximum difference in plate widths for estimation of these parameters. Actual plots of the k_c and k_ϕ pairs support the high correlation of these estimates.

These considerations and the analyses for k_c and k_ϕ made it seem desirable to return to an analysis of the 84 p* values which correspond to the 84 n values analyzed earlier. Such an analysis should give a clearer assessment of the length, width, and interaction effects than we have obtained from analysis of the k_c and k_ϕ values. Any effects found in this analysis must then be present in the k_c and k_ϕ values because of the method of derivation of these estimates from the p* values.

The means for the p^* values are presented in Table 2. Table 3 following gives the analysis of variance for the p^* values. From the table of means it is seen that the width-of-plate effect is much larger than the length-of-plate effect. Both effects, however, may be judged significant from the analysis of variance results. It is interesting to note that the length by width interaction mean square is so small; no interaction is indicated. This result is in contrast to the analysis of the n values, Table 4, for which the interaction effect was judged significant.

CONCLUSIONS FOR PART 2. A large series of experiments have been conducted to study the pressure versus sinkage relation and to estimate the parameters in this relation. The experiments were carried out using a relatively homogeneous soil material, dry mason's sand. Bearing plates used comprised two circles of 2 and 4 inch diameters and 12 rectangular plates varying in size from 1 x 4 to 3 x 10 inches (refer Table 1 for list of plate sizes.) Thus, the length over width ratio of the plates ranged from a maximum of 10 to 1 down to 4 to 3, or from long narrow plates to almost square plates. This range of length over width ratios was selected to cover the range from tracked vehicles to wheeled vehicles with tires.

From the estimates of k_c, k_ϕ, and n plus the estimated pressures for one inch sinkage our analyses show that:

(1) The parameter n varies with the width and length of plate. For one inch width of plate the n value was 0.83; at three inch width, the value was 0.68. Between 2" and 3" width there was little change. With length, n varied from 0.78 to 0.70 over lengths of 4" and 10". The decrease was nearly uniform over the 6" interval.

(2) The estimated pressures for one inch of sinkage (the p^* values) show variation with both length and width of plate but no interaction of the factors is indicated. (Refer to Tables 2 and 3.)

(3) The parameter k_c decreases algebraically with increase in plate length. The algebraic change, however, was much greater when estimates were compared from pairing of plates of 1" and 2" widths with estimates obtained by pairing 2" and 3" widths.

(4) The parameter k_ϕ showed little or no response to length of plate but a large response to width of plate.

(5) The estimates of k_c and k_ϕ are highly correlated. This correlation is negative so that when k_c increases in value k_ϕ decreases in value.

From these analyses it appears that the p versus z relation in the general form $p = (k_\phi + k_c/b)z^n$ is inadequate to predict the pressure-sinkage response. Although not pointed out previously in this paper, it should be mentioned that our analyses to date are based only on the experimental results for sinkage in the range 0.6" to 2". Perhaps it should be added that Dr. Bekker has not claimed that his equation would be adequate in the entire ℓ/b region we have studied.

FUTURE WORK. While much has been learned, there is clearly need for the following:

a. Similar laboratory experiments in other soil media.
b. Experiments with greater depth differences in the soil bins to assess the depth effect, if any, on the parameters under homogeneous soil conditions.
c. Revision of the model to take account of the dimensions of the bearing surface.
d. Improvement of the model to cover a wider range in depth of sinkage, say, from at least 0.5 to 5.0 inches.

Table 2

TABLE OF MEANS FOR p* VALUES

(estimated pressure for one inch sinkage)
summarized by plate dimensions (inches)

SIZE OF PLATE	p*	COMBINED MEANS	p*
Circles		all Circles	6.865
2" diameter	5.509		
4" "	8.222		
Rectangles			
1 x 4	5.188	Lengths (over all Widths)	
1 x 6	4.889	4	6.429
1 x 8	4.701	6	6.378
1 x 10	4.737	8	6.296
2 x 4	6.706	10	6.141
2 x 6	6.671		
2 x 8	6.420	Widths (over all Lengths)	
x 10	6.170	1	4.928
3 x 4	7.392	2	6.492
3 x 6	7.576	3	7.512
3 x 8	7.568		
3 x 10	7.515		

Table 3

ANALYSIS OF VARIANCE OF p^* VALUES
(estimated pressure for one inch sinkage).

SOURCE OF VARIATION	DEGREES OF FREEDOM	MEAN SQUARE
Plates	(13)	8.4927
Circles vs Rectangles	1	2.5279
Between Circles	1	22.0784
Among Rectangles	(11)	7.6181
Widths	2	40.9546
Lengths	3	0.4964
Length by Width	6	0.0668
Plates by Depth	(13)	0.2290
Experimental error	52	0.1793

Table 4

ANALYSIS OF VARIANCE OF n VALUES

(estimated from six replicates grouped together for fitting the pressure-sinkage equation in double-log form)

SOURCE OF VARIATION	Degrees of FREEDOM	MEAN SQUARE
Plates (13)		
Circles vs Rectangles	1	0.0112
Among Circles	1	0.1248
Among Rectangles (11)		
Widths	2	0.1113
Lengths	3	0.0228
Lengths by Widths	6	0.0193
Plates by Depth	13	0.0010
Experimental Error	52	0.0037

REFERENCES

1. Karafiath, L., "Classification of Soils and the Significance of Their Parameters in the Theory of Land Locomotion," Research Report No. 5, 1958, Land Locomotion Research Laboratory, OTAC, Centerline, Mich.

2. Bekker, M. G., "A Practical Outline of the Mechanics of Automotive Land Locomotion," 1955, Land Locomotion Research Laboratory, Detroit Arsenal, Centerline, Mich.

3. Bekker, M. G., "Theory of Land Locomotion, The Mechanics of Vehicle Mobility," 1956, The University of Michigan Press, Ann Arbor, Mich.

4. Taylor, D. W., Fundamentals of Soil Mechanics, John Wiley and Sons, Inc., New York, 1948.

5. Lindley, D. V., "Regression Lines and the Linear Functional Relationship," Jour. of the Royal Stat. Soc., Supplement, 9, 218-244, 1947.

6. Berkson, Joseph, "Are There Two Regressions," Jour. Amer. Stat. Assn., 45, 164-180, 1950.

7. Mood, A. M., Introduction to the Theory of Statistics, 1950, McGraw-Hill, New York.

8. Scheffe, H., The Analysis of Variance, 1959, J. Wiley and Sons, Inc., New York.

EFFECTIVENESS OF CERTAIN EXPERIMENTAL PLANS UTILIZED IN SENSORY EVALUATIONS

J. Wayne Hamman and Jan Eindhoven
Armed Forces Food and Container Institute
Chicago, Illinois

First, I would like to present some specific purposes of sensory testing at the Armed Forces Food and Container Institute. This will be followed by a discussion of the experimental results obtained from the sensory evaluation of four meat products.

SOME PURPOSES OF SENSORY TESTING. You are aware of the numerous food items developed, purchased, stored and consumed by the Armed Forces. A continuous program exists at the Institute to determine whether or not differences in quality or stability exist between different samples of food. Here are some of the most common requirements for conducting these sensory tests:

1. Pre-award evaluation for intent to purchase. When a certain food item (such as peanut butter) is required by the Army, it advertises for bids from manufacturers. Those manufacturers who are interested submit samples of their products for preference evaluation. These samples are taste-tested, and those that are reliably poorer than our standard products are rejected. In this way, sensory testing screens out lower quality products that are relatively unacceptable to the soldier-consumer.

2. Storage stability. Since foods may not be used for several years after they are packed, a considerable amount of research is devoted to extend the shelf-life of a food. Sensory tests are concerned with the preference or intensity of off-flavor changes that take place during storage. Sensory tests are made on foods stored at different temperatures over time up to two years, and more.

3. Packaging studies. Often, a flexible package may be desired for use in the field. However, the relative storage life of food in such a plastic package must be considered if a change is to be made from a canned food.

4. Processing variables. New processing and preservation methods of foods, such as freeze-dehydration of meats, offer new problems in flavor and texture for evaluation. It must be determined whether or not this new product is as desirable as the existing food prepared by other methods.

5. Special Purpose Foods. Recent evaluations have included novel preparations of foods designed specifically for space flight. For example, meat dishes that may be consumed through a straw. Also the new Quick-Serve Meals have been developed which consist largely of pre-cooked dehydrated foods. Preliminary testing is done first at the Institute to determine whether or not these foods are satisfactory enough for further testing in the field among astronauts and soldiers.

SENSORY EVALUATION LABORATORY. In the sensory evaluation laboratory careful attention is given to assure that each sample of food is treated in the same way as every other one in an evaluation. Some of the procedures followed include

1. The random assignment of code numbers to the samples so that subjects will not be biased.

2. In order for the individual to regain sensitivity, that is to get the flavors of a previous sample out of his system, a 30-second time interval is specified between the time that a subject returns the rating of his previous sample and when he receives his next one. Automatic timers are used.

3. In a sensory test each sample is served first, second, third, etc., an equal number of times to minimize position effect. When the number of subjects permits, all possible serving sequences of samples are used. to reduce both serving position and sequence biases that might exist.

4. The number of samples that a subject receives is normally limited to four in order to minimize effects of fatigue and to maintain interest in the evaluation.

PURPOSE OF EXPERIMENT. This experiment is concerned with the effect on sensory results when meat samples are presented to subjects in different combinations and sequences. Specific topics considered are

1. Sequence effects. How is the rating of a sample influenced by the quality of a preceding sample? It is hypothesized that when more highly preferred samples precede those of relatively low preference, the difference between them is emphasized. It is further hypothesized that when the more highly preferred samples follow those of relatively low preference, the difference between them is reduced.

2. Position effects. How is the rating of a sample influenced by the number of preceding samples he has evaluated?

3. Matnitude of error term. How is the size of the error term affected by the quality and sequences of the samples presented?

MATERIALS AND METHODS. Four meat products were evaluated in this experiment: ham, pork, chicken (white) and chicken (dark). Four samples of each of the meat products consisted of two control and two treated samples. An additional preparation variable was included for each meat product which causes both the two control samples and the two treated samples to be considered as non-duplicates. However, the determination of the effect of the additional variable is not the intent of this experiment and will not be specifically considered in this paper.

The subjects sat in a semi-enclosed testing booth for privacy in making evaluations. Each individual received four samples of one of these meat products. These samples were presented one at a time through a turn-table in a wall separating the booth from the serving and preparation area. The subject was asked to state his preference for each sample on a nine-point rating scale. The terminology on this hedonic scale[5] ranged from dislike extremely, coded 1, to like extremely, coded 9, and is shown on the illustrated EAM card (Figure 1) [Figures and Tables can be found at the end of this article] which is used for rating and mechanical data reduction.

Subjects were selected at random from a pool of about 450 employees.

EXPERIMENTAL PLANS. Five experimental plans are considered:

(1) 4! : Conventional plan with all sequences of serving orders. Twenty-four subjects are required for this plan in order to encompass all sequences.

(2) cccc: Two control samples balanced over four serving positions. The two control samples, you will recall, are the non-treated samples and differ by a preparation variable.

(3) tttt: Two treated samples balanced over four serving positions.

(4) cctt: Two control samples followed by two treated samples. The two control samples were served equally often in positions 1 and 2; the two corresponding treated samples were served equally often in positions 3 and 4.

(5) ttcc: Two treated samples followed by two control samples. The two treated samples were served equally often in positions 1 and 2; the two corresponding control samples were served equally often in positions 3 and 4.

Twenty-four subjects were selected at random for each of these plans; all plans were carried out for each of the four products.

EXPERIMENTAL DESIGN. A latin square experimental design[1] was utilized in this study in order to

a. determine the effect, if any, of the sequence of the presentation on the rating for a sample, and

b. reduce the experimental error, if a position effect was present.

Six replications of a 4x4 latin square design were utilized in each plan.

Figure 2 illustrates the allocation of samples for the 4! conventional plan which has all possible sequences. Subjects 1, 2, 3, 4, constitute the first replication, then, and the four samples A, B, C, D, all occur once in each order in a replicate and a subject receives all four samples. The basic Analysis of Variance components, before isolating certain individual degrees of freedom is also given as a part of Figure 2. Since the subjects in a replicate were selected at random, no difference was anticipated between replicates. In the Analysis of Variance treatment x replication and order x replication interactions were pooled into the error term, since there is no reason to expect that these interactions are real.

SEQUENCE RESULTS. Results for the conventional plan (4!) which had all serving orders are shown in Table 1. Mean preference scores for controls were higher than treated for all four of the meat products. The mean differences ranged from 0.56 to 0.79 scale points and significance values were no larger than P = .06 in testing the null hypothesis, namely, that the control mean and treated mean are the same.

Table 2, which presents results for the plan with two control samples served first followed by two treated samples, shows a substantial increase in the discrimination between control and treated samples. The

combined mean difference increased from 0.69 for the conventional plan to 1.00 for this cctt plan and individual probability values declined with the exception of pork which remained about the same (P = .01 vs .02). This phenomenon has been described in previous studies.

In a study with soups and beverages [4] the situation of poor samples following good ones was termed "contrast", that is, one of emphasizing differences. An explanation hypothesized for the phenomenon of "contrast" was that the positive qualities of the good sample are either noticed to be absent or bad qualities are noticed as present in the poor one, thus emphasizing in either case the short-comings of the less preferred one.

Results of the alternative situation where relatively poor samples precede the good ones are given in Table 3. You will notice that the direction of the differences determined in the conventional 4! and contrast plans are not found here. The combined treated samples were rated 0.06 scale points higher than the control in this ttcc plan and none of the individual product differences were statistically significant. This situation where poor samples preceded good ones was termed "convergence"[4], although with these liquids, convergence effects were not found to be statistically significant. An explanation hypothesized is that the presentation of a "poor" sample increases an individual's awareness of the presence of some of the negative characteristics in a "good" sample[4].

Conclusions drawn from these results might be modified somewhat, after a consideration of the position of presentation. We might ask the question: what part of the observed contrast and convergence effects might be due to the fact that samples were presented in positions 3 and 4? We will now proceed to an examination of the positional effect of presentation.

<u>POSITIONAL RESULTS.</u> An examination of the effect of the order in which the sample was received on its rating has been made, considering all five plans. Mean scores by position are given in Table 4. Combined positional means are given in the lower part of this Table for each type of plan. Hypothesis tested were

$$H_1 \quad : \bar{x}_1 > \bar{x}_2$$

$$H_2 \quad : \bar{x}_3 > \bar{x}_4$$

$$H_3 \quad : \bar{x}_{1+2} > \bar{x}_{3+4}$$

Since it was theorized that the position effect (if one existed) may be dependent upon the quality of the product, data were analyzed separately for the control (cc), treated (tt) and the mixed sequence (ct or tc). These are the breakdowns, then, in Table 5. The orthogonal comparisons are shown here for positions 1 vs 2, 3 vs 4 and 1+2 vs 3+4, relating to the mean scores of Table 4.

Probabilistic results from these similar experiments of ham, pork, chicken (white), and chicken (dark) were combined[3] in order to strengthen evidence concerning an effect of position.

Combined evidence for the control (cc) type pairing demonstrated a decrease in preference from position 3 to 4 ($P = .06$) and from the first two positions to the second two positions ($P = .05$). Evidence was not conclusive concerning the decrease in preference exibited from position 1 to 2 ($P = .12$). In the latter case the decrease was 0.22 scale points and was in the hypothesized direction.

Combined evidence for the treated (tt) type pairing did not demonstrate significant positional effects. In the all treated plan the mean for the first two positions was 6.74 contrasted with 6.70 for the last two positions which is reflected by chi-square [6] probability of 0.48. Also the differences between mean preferences, regarding positions 1 vs 2 and 3 vs 4, were not statistically significant.

In the mixed pairing (ct) position 1 was shown to be significantly higher than position 2 ($P = .03$) and the first two positions significantly higher than the last two positions ($P = .03$). Little difference was evidenced between means for positions 3 and 4, however.

A taste testing experiment was reported previously[7] which studied the effect of fatigue over a series of eight samples presented in one sitting. The two foods considered were canned sauerkraut and canned bread with margarine. In a comparison of serving positions 1 or 3 with 5 or 8,

there was found to be no significant difference in preference rating due to the position of the test food, whether it appeared early or late in the eight sample series. These results on the treated samples (no significant difference between positions) would seem to bear out those found on sauerkraut and margarine, since these food items are all relatively low preference items, particularly, the margarine of ten years ago.

Credulence is lent to the theory, then, that while a position effect does appear to exist, the quality of the sample determines whether or not there is a decline in preference with the sequence of presentation.

MAGNITUDE OF ERROR TERMS. A comparison of the magnitude of error terms for the different plans is presented in Table 6. In an analysis of Variance of these variances followed by a Duncan Multiple Range[2] test of means it was determined that the variance for the all control plan was significantly ($P = .05$) smaller than the three plans having both control and treated samples. Also, the error variance of the all treated plan was significantly ($P = .05$) smaller than for the conventional 4! and the plan with two control samples followed by two treated samples (cctt).

These results are in line with anticipations, however, since the ranges of ratings of individuals within the all treated and all control plans are less than for the other plans. Hence, the magnitude of disagreement in preference would be expected to be less for these plans. The differences in magnitude of these variances, though, do point out the necessity for analyzing differences between means, such as those for order, individually, for each plan.

SUMMARY. The preference rating scale is normally used for comparative purposes between samples. However, one can see the effect that might occur on the magnitude of scores of experimental samples, depending upon the quality of standards or controls with which they are compared.

Also the sequence in which relatively good and poor samples were presented made a considerable difference in what conclusions would be drawn. Evidence concerning the effect of the position of presentation was given. For the higher preference samples it was demonstrated that a fatigue effect or sensitivity effect was present causing a decline in the rating of subsequent samples (although not so conclusive between positions 1 and 2 where $P = .12$). For the lower preference samples there was not determined to be a decline in ratings in subsequent positions.

If one desired to minimize Type II statistical error (acceptance of the null hypothesis when it is false), then, one would wish to present test samples after the standards. If the test samples were poorer, the difference would be emphasized by a contrast effect pointed out earlier. The statistical soundness of such a procedure might be questioned, however, since the magnitude of the contrast effect would be expected to be greater when the difference between samples was greater. This would affect the probability statements concerning a "true" difference in the population to a corresponding unknown degree.

REFERENCES

(1) Cochran, G. C. and Cox, G. M. Experimental Designs, John Wiley and Sons, New York, 1950, pp. 103-112.

(2) Duncan, David B. Multiple Range and Multiple F Tests, Biometrics, March 1955, pp. 1-42.

(3) Jones, L. V. and Fiske, D. W. Models for Testing the Significance of Combined Results, The Psychological Bulletin, Vol. 50, No. 5, pp. 375-382.

(4) Kamenetsky, Joe, Contrast and Convergence Effects in Ratings of Foods, Journal of Applied Psychology, Vol. 43, No. 1, 1959.

(5) Peryam, David R., et. al. Food Preferences of Men in the U. S. Armed Forces. Department of the Army, Quartermaster Food and Container Institute for the Armed Forces, January 1960.

(6) Snedcor, G. W. Statistical Methods, The Iowa State College Press, Ames, Iowa, 1946.

(7) Symposium sponsored by the Quartermaster Food and Container Institute for the Armed Forces, Food Acceptance Testing Methodology, October 1953, pp. 92-99.

Figure 1. Preference Rating Card Used for Sensory Evaluations

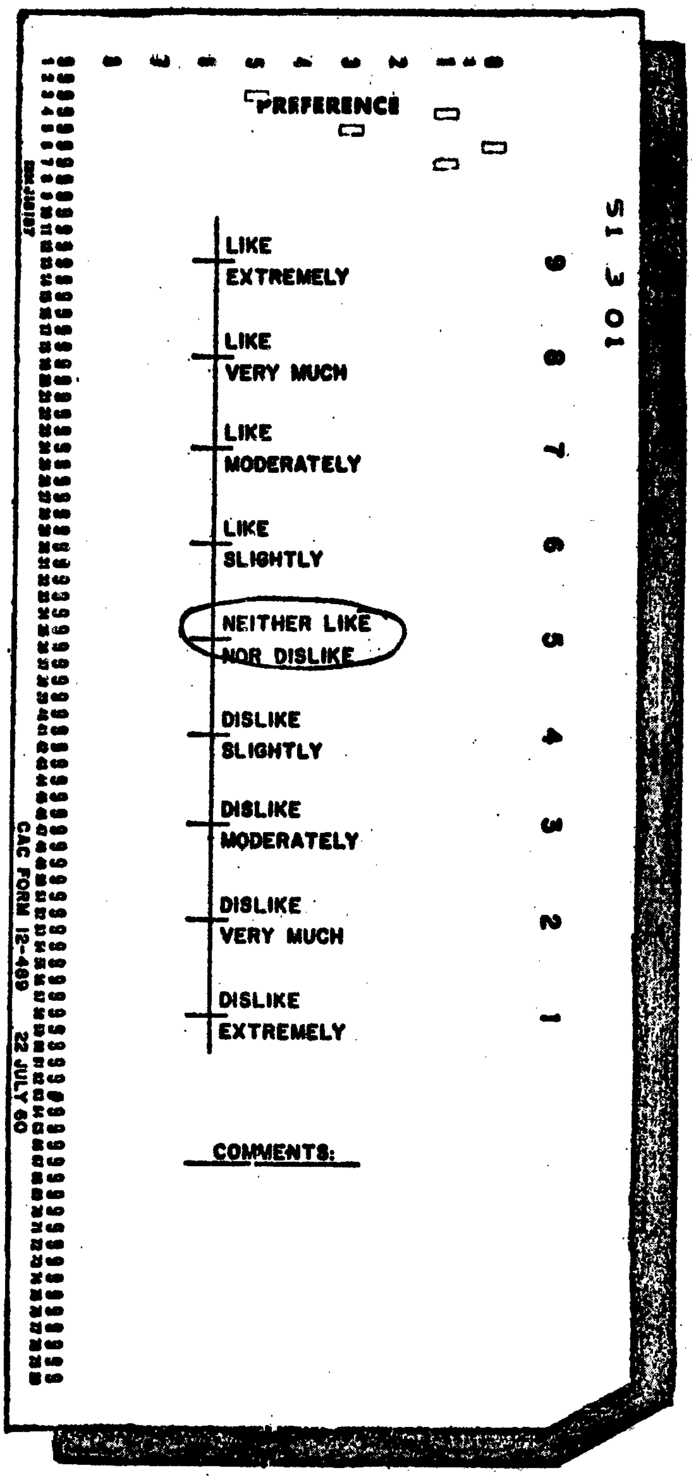


FIGURE 2. Allocation of Samples for the 4! Experimental Plan Utilizing a Latin Square Design

Subject	Order 1	Order 2	Order 3	Order 4
1, 5, 9, 13, 17, 21	A	B	C	D
2, 6, 10, 14, 18, 22	D	A	B	C
3, 7, 11, 15, 19, 23	C	D	A	B
4, 8, 12, 16, 20, 24	B	C	D	A

Analysis of Variance Componenets

Source of Variation	Degrees of Freedom
Subjects	23
Orders	3
Treatments	3
Error	66
Total	95

TABLE 1. Mean Preference Ratings of Four Meat Samples in All Possible Serving Orders, Termed a 4! Plan*

Food	Control	Treated	Control - Treated Diff.	Signif. Level for Diff.
Ham	7.17	6.46	0.71	.04
Pork	6.56	5.77	0.79	.01
Chicken, W.	7.33	6.77	0.56	.02
Chicken, D.	6.75	6.06	0.69	.06
Combined			0.69	

TABLE 2. Mean Preference Ratings of Four Meat Samples Where Two Control Samples Were Presented, Followed By Two Treated Samples, Termed a cctt Plan*

Food	1st & 2nd Position Control	3rd & 4th Position Treated	Control - Treated Diff.	Signif. Level for Diff.
Ham	7.40	5.75	1.65	.001
Pork	6.52	5.88	0.64	.02
Chicken, W.	7.14	6.35	0.79	.005
Chicken, D.	6.71	5.81	0.90	.005
Combined			1.00	

TABLE 3. Mean Preference Ratings of Four Meat Samples Where Two Treated Samples Were Presented, Followed By Two Control Samples, Termed an ttcc Plan*

Food	1st & 2nd Position Treated	3rd & 4th Position Control	Control - Treated Diff.	Signif. Level for Diff.
Ham	6.85	6.60	-0.25	n.s.
Pork	5.96	5.92	-0.04	n.s.
Chicken, W.	6.90	7.04	0.14	n.s.
Chicken, D.	5.90	5.83	-0.07	n.s.
Combined			-0.06	

*Individual Means represent ratings of 2 samples by 24 subjects.

TABLE 4. Mean Preference Ratings for Each Serving Order of Four Meat Products for Five Experimental Plans. Individual Means Represent 24 Subjects.

Food	Plan/	Order of Presentation 1st	2nd	3rd	4th
Ham	4!	7.71	6.87	6.21	6.46
	cccc	7.46	7.42	7.37	7.21
	tttt	7.12	7.21	6.83	6.96
	cctt	7.71	7.08	5.50	6.00
	ttcc	6.54	7.17	6.62	6.58
Pork	4!	6.25	5.88	6.38	6.17
	cccc	6.62	6.54	6.79	6.88
	tttt	6.50	6.25	6.21	6.50
	cctt	6.58	6.47	5.75	6.00
	ttcc	5.75	6.17	6.17	5.67
Chicken, W.	4!	7.20	7.04	6.71	7.25
	cccc	7.58	7.46	7.29	6.92
	tttt	6.88	7.04	7.00	6.92
	cctt	7.42	6.88	6.38	6.33
	ttcc	6.83	6.96	7.17	6.92
Chicken, D.	4!	7.12	6.12	6.38	6.00
	cccc	6.83	6.50	6.71	6.33
	tttt	6.50	6.46	6.83	6.33
	cctt	6.67	6.75	5.58	6.04
	ttcc	6.04	5.75	6.08	5.58
Combined	4!	7.07	6.48	6.42	6.47
	cccc	7.12	6.98	7.04	6.84
	tttt	6.75	6.74	6.72	6.68
	cctt	7.10	6.80	5.80	6.09
	ttcc	6.29	6.51	6.51	6.19

TABLE 5. Probability Values* for Orthogonal Comparisons of Position Effect of Mean Preference Ratings in Which Four Meat Samples Were Presented to a Subject

Position	Meat /	Pairing	cc		tt		ct
		Plan	4	2	5	3	1
1 vs 2	Ham		.11	.44	.95	.60	.19
	Pork		.38	.39	.82	.21	.18
	Chicken, W.		.07	.32	.66	.73	.29
	Chicken, D.		.58	.12	.21	.45	.02
	χ^2 Comb'd P		.13	.27	.81	.61	.03
			(.12)		(.82)		
		Plan	5	2	4	3	1
3 vs 4	Ham		.44	.28	.83	.64	.70
	Pork		.14	.62	.75	.82	.25
	Chicken, W.		.22	.08	.44	.38	.96
	Chicken, D.		.09	.09	.87	.08	.22
	χ^2 Comb'd P		.10	.10	.93	.41	.58
			(.05)		(.79)		
		Plan		2		3	1
1+2 vs 3+4	Ham			.22		.15	.008
	Pork			.88		.45	.77
	Chicken, W.			.013		.50	.26
	Chicken, D.			.22		.66	.11
	χ^2 Comb'd P			.06		.48	.03

*Probability Values Presented in This Table Reflect the Test of Significance Regarding the Hypotheses of $\bar{x}_1 > \bar{x}_2$, $\bar{x}_3 > \bar{x}_4$ and $\bar{x}_{1+2} > \bar{x}_{3+4}$.

Combined Results Were Obtained by the Chi-Square Method of Combining Results of Similar Experiments(3). Individual Probabilities in the Table are Based on an N of 24.

TABLE 6. Error Terms from Five Experimental Plans and Four Meat Products

Plan /	Ham	Pork	Chicken, W.	Chicken, D.	Composite*
41	2.7386	1.9697	1.1203	2.9246	2.1883
cctt	3.0850	1.6721	1.4882	1.9670	2.0531
ttcc	1.6356	2.5553	1.1933	1.5490	1.7333
tttt	1.6360	1.1497	0.8085	1.4420	1.2590
cccc	0.9134	1.0659	0.7993	0.9171	0.9239

*Bracketed numbers indicate the variances which are not significantly different at the probability level of 0.05.

AN EVALUATION OF RADIATION-PROCESSED FOODS FOR MILITARY RATIONS*

Donald M. Boyd
Research Analysis Corporation
Bethesda, Maryland

In the fall of 1960 a research group from our organization (then known as ORO) was asked to investigate the possible operational, logistical, and economic advantages to the armed forces of employing radiation-processed foods in the military feeding system and to provide a basis to assist the Army in making decisions on the irradiated-food research program.

The preservation of food by sterilization with ionizing radiation is a relatively new concept and at that time had not been attempted commercially. Experimentally, many foods have been irradiated to determine the value, safety, and efficacy of such processing. Various radiation doses have been employed under different conditions of exposure and of associated treatment techniques. The ultimate goal is the attainment of a process that would safely, and at a reasonable cost, preserve foods so that they could be stored in a fresh-like and wholesome condition for long periods of time without refrigeration. Because meats are the highest-valued items in military rations, special research efforts have been placed on the development of radiation-processed meat items for ration components.

Considerable progress has been made since The Quartermaster General's extensive research program on irradiated foods began in 1953. However, their plans for the construction and operation of a developmental pilot plant were indefinitely suspended on the recommendation of the Director of Research and Development of the Army, in 1959. This action was taken partly because of the uncertainty of the wholesomeness of the foods, but mainly on the basis of need for adequate reliable information concerning the operational, logistical, and economic advantages that would justify the use of irradiated foods in military rations and the construction and operation of a pilot plant. In March 1960, a revised Army program on radiation preservation of foods was approved for wholesomeness studies and fundamental research toward development of

*The data presented in this paper have been published in greater detail as ORO-SP-174, "Radiation-Processed Foods as a Component of the Armed Forces Feeding Systems," August 1961.

end items. The Operations Research study, undertaken by RAC, was included in the plan of that new program.

Our investigation began with an analysis of the technological status of radiation-preservation of foods and the additional research effort needed to attain fully acceptable products for incorporation into military rations during the 1965-1975 time frame. Attention was given to various concepts of military feeding and the tactical operational requirements of various armed-forces units for ration support. Then particular consideration was given to the logistical implications involved in the integration of irradiated components in military rations, including savings in manpower, storage, equipment and supplies. The costs of processing, storing and transporting irradiated foods were concurrently studied and our estimates compared with estimated costs of the freeze-dehydration process, and the costs of the commercial canning and freezing processes.

An examination was also made of the different types of radiation sources that could be used for processing foods, and of the availability, costs, and efficiency of these radiation sources.

As a last step, we investigated the feasibility of establishing a mobilization base containing irradiated meats as ration components, and estimated the number and cost of accelerators needed for a production base.

Our findings were published in August of last year.

In order to be approved by the Food and Drug Administration, irradiated foods for public consumption must not show radioactivity levels that are distinguishable from background. Reports have shown that foods can be processed by gamma radiation from Co^{60} or by electron accelerators below energy levels of 10 Mev without inducing measurable radioactivity in the foods. Radiation preservation of meats requires an exposure dose of about 4.5 megarads preceded by short heating to an internal temperature of about 160° F. Reports showed that beef and pork processed this way had remained acceptable for at least 25 months at 70°F storage temperature and for 16 months at 100°F. Bacon and ham had been stored about one year and chicken for one and one-half years with good acceptability.

We found that numerous, extensive studies to determine the wholesomeness of irradiated foods had been conducted, both in-house and

under contract to the Army Quartermaster Corps and the Surgeon General.

Objective analyses of the results of these studies, including long-term animal feeding tests, showed no harmful effects attributable to radiation processing beyond correctable vitamin loss. However, prudently cautious recommendations by the Surgeon General included a few more years of research for completion of the wholesomeness study program.

In the future feeding concept shown in Fig. 1, single meal modules would be employed as follows:

(a) The 25-in-1 uncooked meal might be used behind the contact area. This ration will contain canned, dehydrated, and irradiated foods, and will be served by trained food service personnel in a unit-mess type of feeding.

(b) From the reserve area forward into the contact area the 25-in-1 precooked quick-serve meal would be used where the tactical situation precludes the preparation and serving of the 25-in-1 uncooked meal. This ration will contain precooked freeze-dehydrated foods as the major component, which will be prepared by one or two individuals by pouring hot water directly into the food packages and serving it on disposable trays packed in the carton with the meal. Trained food service personnel would not ordinarily be involved. Under some circumstances precooked irradiated meats might be used in place of freeze-dehydrated meats.

(c) In situations where small groups are dispersed from their units for long periods, the 6-in-1 precooked ready-to-serve meal would be used. It will contain the same type of foods as the 25-in-1 meal.

(d) In the contact area especially and under certain conditions to the rear, the tactical situation will often require the use of an individual ready-to-eat meal. This will contain precooked irradiated foods that will normally be eaten cold but could be warmed by the individual when his situation will allow it. Flexible packaging will add to the value of this ration.

(e) Individuals would also be issued an individual combat food packet for emergency use. This will be a small, compressed, high-caloric-

content food item totalling about 1000 calories. This is not intended to replace a meal when other rations can be provided but is to be capable of sustaining a man for as long as 2 to 10 days under emergency situations without appreciable loss of efficiency and without irreversible physiological damage.

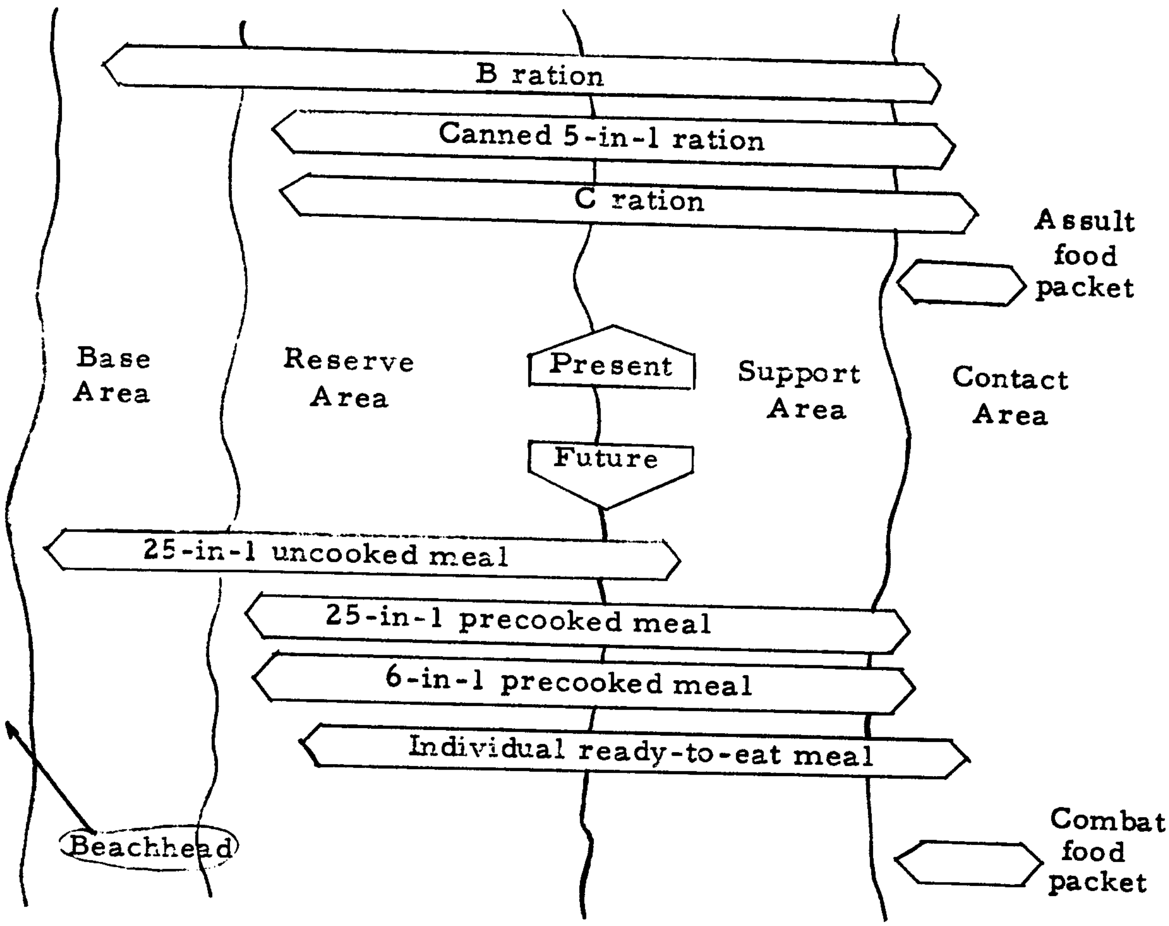


Fig. 1 -- Present and Future Feeding Concepts

With the current emphasis on mobility of US forces vs enemy mass a continuing requirement exists for improved logistical operations to support such mobility. This requirement pervades all classes of supply. Although class I supply involves only a small fraction of the tonnage of class III fuels, simplified rations do contribute to reducing fuel consumption as well as equipment and manpower requirements, and to improving mobility.

In seeking a logistical advantage of employing irradiated rations, we estimated the potential savings these rations would permit if the logistic burden of field kitchens could be eliminated from forward units. An Army of 2 million men was considered, using a distribution of men in a theater based on data from FM 101-10, as shown in Tables I and II.

TABLE I

COMPOSITION OF THEATER SLICE BY ASSIGNMENT

Assignment	Troops
Basic division	13,961
Nondivision	18,540
Theater overhead	24,750
Army	(10,750)
Air Force	(14,000)a
Total	57,251

[a]Two Air Force wings including Army support.

TABLE II

COMPOSITION OF AN OVERSEAS ARMY SLICE

Item	Division	Corps[a]	Army (rounded)[b]
Men			
Armor	14,600	14,600	---
Infantry	13,748	41,244	---
Corps, nondivision	---	74,160[c]	---
Theater overhead	---	99,000[d]	---
Total		229,004	687,000
Field Kitchens			
Armor	98	98	---
Infantry	70	237	---
Total		335	1,005
Bakery companies (mobile)	---	---	5
Refrigeration companies (mobile)	---	---	1

[a]Corps has three infantry divisions and one armored division plus non-division troops.

[b]Army has three corps.

[c]18,540 x 4

[d]24,750 x 4.

In Table III can be seen the requirements for the operation of field kitchens for 2 million men in a theater of operations. Of particular importance is the fuel consumption rate of over 65,000 tons per year.

TABLE III

REQUIREMENTS FOR FIELD KITCHENS
(For 2 million men overseas)

Item	Quantity
No. of field kitchens	2,925[a]
Men	14,625
Trucks	2,925
Water, millions of gal/year	
Equipment	213[c]
Men	80[d]
Total	293
Fuel, tons/year	
Trucks	4,100[e]
Cooking	56,000[f]
Pump water	5,130[g]
Total	65,230

[a]From Table II $(1005 \times (2 \times 10^6))/687{,}000 = 2925$ field kitchens.

[b]Based on five men per field kitchen.

[c]Based on 200 gal/day/kitchen for washing and rinsing mess trays plus all else not consumed directly by men.

[d]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[e]Based on 2000 miles/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb/gal.

[f]Based on 15 gal/day/kitchen.

[g]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

The requirements for mobile bakery companies are shown in Table IV. These companies bake bread for troops in the field, but may also be

assigned to supplement the production of garrison bakery units as the situation demands. During overseas use these bakeries require over 8,000 tons per year of fuel per 2 million men.

TABLE IV

REQUIREMENTS FOR MOBILE BAKERY COMPANIES
(for 2 million men overseas)

Item	Quantity
No. of bakery companies	15[a]
Men	2130[b]
Trucks	315[c]
Water, millions of gal/year	
Equipment	39[d]
Men	12[e]
Total	51
Fuel, tons/year	
Trucks	2200[f]
Ovens	1150[g]
Electric generators	3950[h]
Pump water	890[i]
Total	8190

[a]From Table II $(5 \times (2 \times 10^6))/687{,}000 = 14.6$ bakery companies.

[b]Based on 142 men/company.

[c]Based on 21 trucks/company.

[d]Based on 100 gal/hr/platoon and three platoons per company.

[e]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[f]Based on 10,000 miles/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb/gal.

[g]Based on 10 gal/day/oven and six ovens per company.

[h]Based on three generators (25 Kw each) per company running continuously; specific fuel consumption = 0.6 lb/hp-hr.

[i]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

Mobile refrigeration companies deliver perishable foods from depots to supply points, but many also use their semitrailer vans as fixed refrigerators as the situation demands. During use in the theater of operations, these companies require fuel for gasoline-driven refrigerant compressors and trucks. These requirements are illustrated in Table V.

TABLE V

REQUIREMENTS FOR MOBILE REFRIGERATION COMPANIES
(For 2 million men overseas)

Item	Quantity
No. of refrigeration companies	3[a]
Men	564[b]
Trucks	165[c]
Water for men, millions of gal/year	3[d]
Fuel, tons/year	
Trucks	1,150[e]
Refrigeration equipment	10,600[f]
Pump water	50[g]
Total	11,800

[a]From Table II $(1 \times (2 \times 10^6))/687{,}000 = 2.9$ refrigeration companies.

[b]Based on 188 men/company.

[c]Based on 55 trucks/company, of which 48 are 7 1/2-ton semi-trailers and 7 are 2 1/2-ton trucks.

[d]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[e]Based on 10,000 mile/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb./gal for both truck types.

[f]Based on 5-hp motors, specific fuel consumption of 0.6 lb/hp-hr, and 7000 hr/year of operation.

[g]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

If the use of irradiated foods and freeze-dehydrated foods would permit the elimination of refrigerated warehouses, a substantial saving in the overseas logistical effort shown in Table VI could be attained.

TABLE VI

REQUIREMENTS FOR REFRIGERATED WAREHOUSES
(For 2 million men overseas,
warehouse size, 20 by 100 ft)

Item	Quantity
No. of refrigerated warehouses	173
Men	519[a]
Water, millions of gal/year	
Equipment	6
Men	3[b]
Total	9
Fuel, tons/year	
Electricity generation	12,700
Pump water	100[c]
Total	12,810

[a]Based on three men per warehouse (ORO estimate).

[b]Based on 15 gal/man/day.

[c]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

Table VII summarizes those requirements for kitchens, bakeries, and refrigeration facilities which could be reduced from the total logistical

effort by employment of irradiated foods and dehydrated foods in military rations. Net savings that could be attained in fuel alone are shown in Table VIII. The delivery of bulk fuel by truck is one of the largest problems in theater logistics.

TABLE VII

REQUIREMENTS FOR KITCHENS, BAKERIES, AND REFRIGERATION FACILITIES
(For 2 million men overseas)

Facility	Men	Trucks	Water, millions of gal/year	Fuel tons/year
Field kitchens	14,625	2925	293	65,230
Bakery companies	2,130	315	51	8,190
Refrigeration companies	564	165	3	11,800
Refrigerated warehouses	519	0	9	12,800

TABLE VIII

NET FUEL SAVINGS
(For 2 million men overseas)

Item	Fuels saved, tons/year
Field kitchens	65,230
Bakery companies	8,190
Refrigeration companies	10,600[a]
Refrigerated warehouses	12,700[b]
Total for irradiated foods	96,720[c]
Preparation of freeze-dehydrated food	-5,080[d]
Total for freeze-dehydrated foods	91,640[e]

[a]Refrigerating equipment only.

[b]Refrigerating equipment only. All electricity generation is assumed to be for refrigeration purposes.

[c]Fuel for electricity for truck shops is less than 0.5 percent of this total.

[d]Fuel consumed in heating of water for freeze-dehydrated foods.

[e]Corresponds to 250 tons/day and is not sensitive to the assumed percentage of men who may eat freeze-dehydrated food regularly. For example, 50 percent freeze-dehydrated food in a theater yields 96,720 - 10,160 = 86,560 tons/year, which corresponds to 237 tons/day.

Our cost analysis for processing and transporting irradiated foods showed that these costs were similar and competitive to those of freeze-dehydrated foods and foods preserved by other commercial means.

Our analysis of costs of operation of electron accelerators included the result that the cost of these machines per kilowatt of output power decreases with the increase in the power rating of the accelerator. In addition, at any given radiation dose the output in terms of food processed is proportional. When we applied our operational costs analyses to the problem of establishing a mobilization base, we found that seven 100-Kw accelerators or twelve 30-Kw accelerators would be required to process the total annual meat ration requirements per 1 million men (Fig. 2). The costs of processing this amount of meat is shown in Table IX.

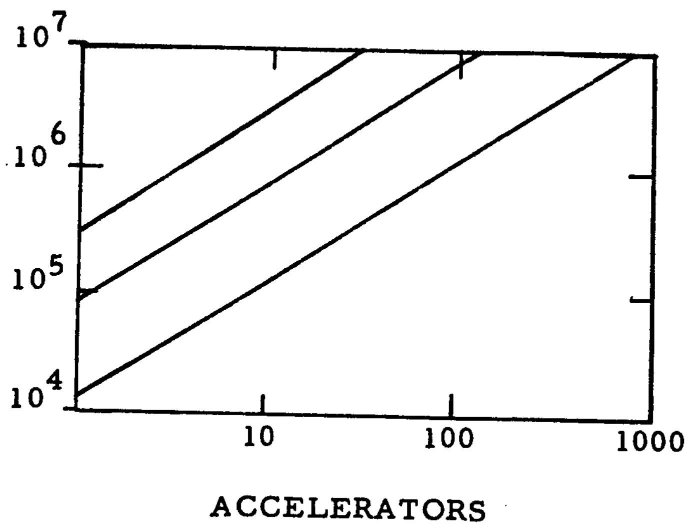


Fig. 2 -- Number of Accelerators Needed to Treat Meats for Armed-Forces Personnel

TABLE IX

ANNUAL CAPITAL AND OPERATING COSTS FOR MEAT PROCESSING
(In millions of dollars)

Manpower level	First Year		Subsequent Years[a]	
	30 Kw	100 Kw	30 Kw	100 Kw
10^4 men	0.039 - 0.057	0.016 - 0.029	0.020	0.010
10^6 men	3.9 - 5.7	1.6 - 2.9	2.0	1.0
10^7 men	39.0 - 57.0	16.0 - 29.0	20.0	10.0

[a]No amortization, operating costs only.

The QMC generally stocks meat reserves in a 15-month supply (composed of a 12-month operational reserve plus a 3-month safety reserve). We determined that a mobilization base reserve of 15 months' rations containing irradiated meat components could be obtained by establishing a production base and operating for one year nine 100-Kw electron accelerators per million men supplied.

About 2 years of total lead time would be required to establish the production base and produce the 15-month reserve supply.

It is likely that only 50% of the meat components of future rations in this system would contain irradiated meats; the rest being processed by other means. Under this consideration the electron accelerators required for the reserve supply would be reduced to 5.

The results of our studies permitted us to make certain conclusions, some of which were as follows:

CONCLUSIONS.

1. The use of rations containing irradiated foods instead of B and C rations and the elimination of field kitchens in general war could result in logistical savings equivalent to 97,000 tons of fuel/year/2 million men in the theater of operations.

2. The logistical savings gained by employing only dehydrated foods instead of B and C rations would be equivalent to 91,000 tons of fuel/year/2 million men.

3. In 1965-1975 irradiated foods could have a distinctive advantage over all other types of foods in providing an operationally suitable individual combat meal that would be well received by fighting men.

4. The estimated cost of radiation processing of foods would be competitive with the costs of the thermal-canning, freezing, and freeze-dehydration processes.

5. About 2 years would be required to obtain a mobilization base composed of a 15-month reserve supply of rations with irradiated meats comprising 50 percent of the meat components. This time includes the estimated 9 to 12 months required to establish radiation facilities and the 12 months required to process the rations.

6. To process the 15-month supply of rations, five 100-Kw accelerators/million men would be required at a cost of $1 million to $1.8 million.

A CRITIQUE OF THE EVIDENCE RELATING DIET AND CORONARY HEART DISEASE

George V. Mann
Division of Nutrition, Vanderbilt University
School of Medicine
Nashville, Tennessee

It may be useful for me to review the problem of coronary heart disease (CHD) from the special viewpoint of a nutritionist. While this view may have some prejudice it seems relevant because of the frequent association of diet with CHD and the widespread lay interest in the problem.

Coronary heart disease may seem to have risen like an epidemic among us. It is a complicated task to determine whether this rise to prominence is real or only made apparent by changing techniques. It would be an interesting task for someone to relate the time course of the prevalence of CHD to the marketing of electrocardiographs. To my knowledge this has not been done. One might have expected a rise of CHD when the ECG became available for diagnosis. Dr. Lew of the Metropolitan Life Insurance Company has shown a remarkable explanation for the distribution by states of coronary heart disease in the United States (1)(2), (Fig. I and II). It must be clear that we see what we look for.

A more subtle influence is that of "competing causes" (3). Even when age specific rates are considered we may be baffled in understanding the entire effect of the removal of diseases which typically kill at an earlier age than does coronary heart disease.

The errors of reporting cause of death are well known (4). It is less than even money that an autopsy will confirm the clinical impression and only a small proportion of all deaths are followed by necropsy. Since reporting causes of death, like ladies hats, tends to change with fashion it is easily possible for the mortality rates to be strongly influenced by the current fashion and this is conveniently done since the selection of the first cause in the presence of multiple causes of death will determine the final tabulations.

Finally we must concede that it is possible that an apparent rise of prevalence of coronary heart disease is real and that this is a reflection

of the introduction of a new and potent causal factor that we must identify and adjust in order to control CHD.

The interest of nutritionists in this problem like their interest in most diseases stems from the ancient judgement that a man may be sick because of "something he et". This explanation has proved so attractive that we have a second epidemic, a scourge of nutritionists. These newcomers, coupled with the food industries, have made food and feeding a highly complicated and even dangerous business.

The essential series of hypotheses upon which most research is presently based may be shown as follows:

Diet $\overset{1}{}$ Hypercholesteremia $\overset{2}{}$ Atherosclerosis $\overset{3}{}$ Clinical Events

The evidence to support the first relationship is at best indecisive. The question was brought to prominence by A. Keys (5) who based this contention on a curious selection of food-mortality data of the World Health Organization (Fig. III). Aside from the fact that the hypothesis is based on tenuous population data that might as easily be explained in other ways (6) it has proven impossible to show in retrospective studies that persons with CHD eat differently than those without (Fig. IV).

The dietary behavior of 983 persons in the Framingham Study has been measured by Georgiana Pearson in the past four years (7)(8)(9). The reproducibility of the method whether by one person (Fig. V) or by a second observer (Fig. VI) is good. We are confident that these people were well classified but we can find no relationship between either cholesterol level (Table 1) or experience with CHD and the way these people eat. Morris, Marr and Heady (10) have found no diet-cholesterol disease relationship in their population of bank clerks (11). Their method of measuring diet does not reproduce quite as well as ours. (Table 2).

The entire problem is complicated by the prevailing imprecision of the measurement of cholesterol. Consider, for example, the data of Rivin (12) (Table 3) who compares hospital and commercial laboratories. We have compared several methods applied in a research setting (13) (Table 4). If one adds to this technical variation the considerable biological variation of serum cholesterol with time (14) it is clear that the

central element of the hypothesis may be so badly estimated that this disqualifies our most convenient index (Table 5).

We are at least as bad off in measuring atherosclerosis, the anatomical lesion we believe to be the basis for the clinical disorders. We cannot visualize these lesions in life and even after death to do more than make qualitative descriptions is difficult. You can appreciate that an element of probability determines whether the plaque is critically placed in the cardiovascular system.

The clinical manifestations of CHD are varied (Table 6). A disturbing number, disturbing at least for the biometrician, are completely occult events called "silent coronaries" because they do not cause important clinical signs. The cerebral events, strokes, are even more obscure because we have less precise ways to determine and localize these, having no equivalent for the ECG.

There are several prospective dietary studies under way which propose to change the experience with CHD by altering the diet. The dietary regimens of some of these are summarized in Table 7. The most ambitious of these called the National Diet Heart Disease Study is directed by Dr. Irvine Page and sponsored by the National Heart Institute (15). It is now in the feasibility phase, that is, the determination of whether families can be recruited, supplied with suitable food and kept under surveillance while consuming the diet for the measurement of cholesteremia and the evaluation of cardiovascular disease status. If proven feasible, this experiment will be extended to larger numbers in order to answer the critical question--will dietary changes modify the course of CHD?

The smaller trials of diet, for example, that of Dayton at Los Angeles (16) and Rinzler with the Anti-Coronary Club (17) in New York have usually obtained about a 15% reduction of serum cholesterol in the best circumstances, that is, when the starting level is high. However, many subjects who do follow the diet do not respond and some who respond initially drift back up with time. We must conclude that dietary treatment, if effective, is a relatively impotent agent. We must conclude also that diet has been overemphasized as a cause of CHD and that dietary modifications are proving relatively ineffectual control measures.

Table 1

DIETARY INTAKES - AMERICANS 1957-58

FRAMINGHAM HEART STUDY

ARRANGED BY SERUM CHOLESTEROL LEVEL

Men	N	Calories	Fat g.	Protein g.	Chol. mg.
High Cholesterol	17	3127	149	113	703
Low Cholesterol	39	3487	163	126	721
Random Sample	133	3333	157	122	735

Table 2

COMPARISON OF REPEATABILITY FOR 2 METHODS OF MEASURING DIET

Nutrient	Heady - Bank Clerks 1 week's weighed intake	Framingham research diet history
	r - consecutive weeks	r - 2 year interval
Calories	0.80	0.92
Protein (gm)	0.67	0.72
COH (gm)	0.84	0.90
Fat (gm)	0.79	0.88

Table 3

Cholesterol Measurement

Rivin, et al., J. A. M. A., 166:2108, 1959

Values in Mgm%

Serum	Author	V. A.	Univ.	Commercial 1.	Commercial 2.	Commercial 3.
A	529	479	480	598	513	411
	487	418	541	500	451	318
B	260	240	255	291	273	183
	273	233	296	263	272	191
C	218	213	275	312	255	180
	249	220	288	252	246	172
Method	K-S	K-S	B1	B1	PSG	Sheftel

Table 4

Evaluation of Methods for Serum Cholesterol

		Methods							
		Abell		FeSac		Pearson		Sackett	
Level (mgm%)	N	$\overline{X}$	T. E.	$\overline{X}$	T. E.	$\overline{X}$	T. E.	$\overline{X}$	T. E.
<210	15	188	6.4	180	5.6	179	7.7	237	3.2
211-274	15	232	4.1	228	5.7	225	8.8	288	3.2
275-499	14	368	5.2	384	7.6	340	15.1	452	12.9
>499	15	672	7.7	667	10.4	643	26.9	837	14.2
all levels	59	365	6.0	365	7.6	347	17.1	454	14.2

Table 5

Serum Cholesterol Variation

68 men - measured twice weekly - 10 weeks

S_T = total variation $\quad$ S_E = laboratory variation $\quad$ S_B = biological variable

where $\quad NS_B^2 = NS_T^2 - 1/2\ NS_E^2$

$$\overline{X}S_T = 20$$

$$\overline{X}S_E = 7$$

$$\overline{X}S_B = 13$$

Then: For 95% assurance of effect 2 x 20 = <u>40</u> <u>mgm</u> % <u>minimum change</u>.

Table 6

THE MANIFESTATIONS OF CORONARY HEART DISEASE

Of 100 Men with "Events"

30 drop dead

20 are "silent"

10 die a little later

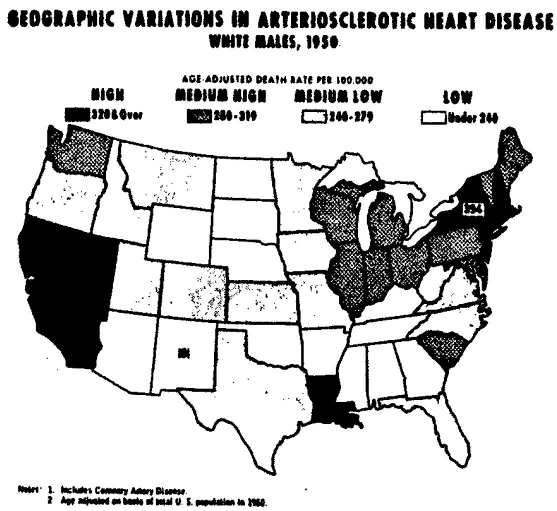
40 recover

Table 7

DIETARY REGIMENS -- OBSERVED AND FAT RESTRICTED

	Framingham (Men)	Page	Rinzler	Dayton	American Heart
Calories	3075	2000	2400	2430	2800
Protein (gm)	112	70	140	94	85
Fat (gm)	154	90	81	106	75
% Cal.	45	41	32	40	36
Cholesterol (mgm)	705	<200	200	380	200
PUS/S	0.3	1.5	1.0	1.7	1.1

Figure I



Data of Enterline and Stewart, Reference 1.

Figure II

CORRELATION BETWEEN MORTALITY FROM ARTERIOSCLEROTIC HEART DISEASE AND INTERNISTS PER 100,000 WHITE PERSONS

United States 1950

Region	Age-adjusted Death Rate per 100,000	Number of Internists** per 100,000
Middle Atlantic	273	12.6
New England	250	10.5
Pacific Coast	233	9.7
East North Central	214	7.3
South Atlantic*	193	7.6
West North Central	183	6.7
Mountain	179	6.5
West South Central	176	5.8
East South Central	160	4.5

* Excludes District of Columbia
** Includes cardiologists

Includes Coronary Heart Disease. Death rates age adjusted on basis of total U. S. population in 1950.

This material was published by E. A. Lew - Reference 2.

Figure III

DIET AND MORTALITY FROM HEART DISEASE IN
22 COUNTRIES 1951-53
MEN 55-59 YRS.

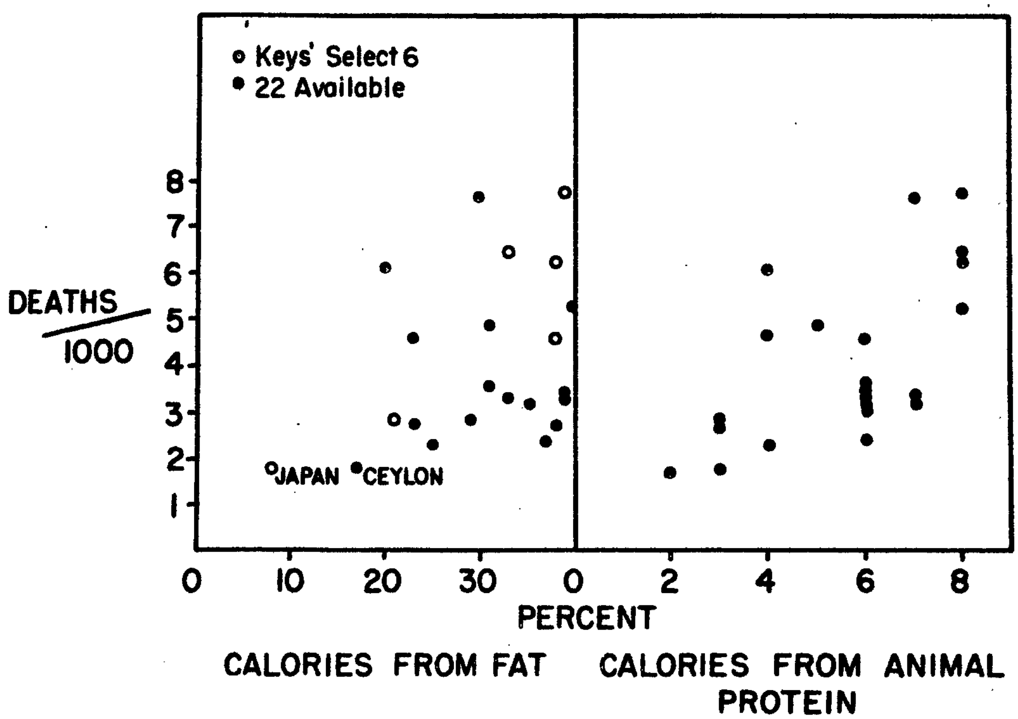


Figure IV

OBSERVATION OF DIET PATTERN AND EXPERIENCE WITH CORONARY HEART DISEASE

Observer		Association Observed
Wilkinson	Michigan	no
Rosenman	California	no
Zukel	North Dakota	no
Mann	Massachusetts	no
Morris	London	no
Keys	Minnesota	?

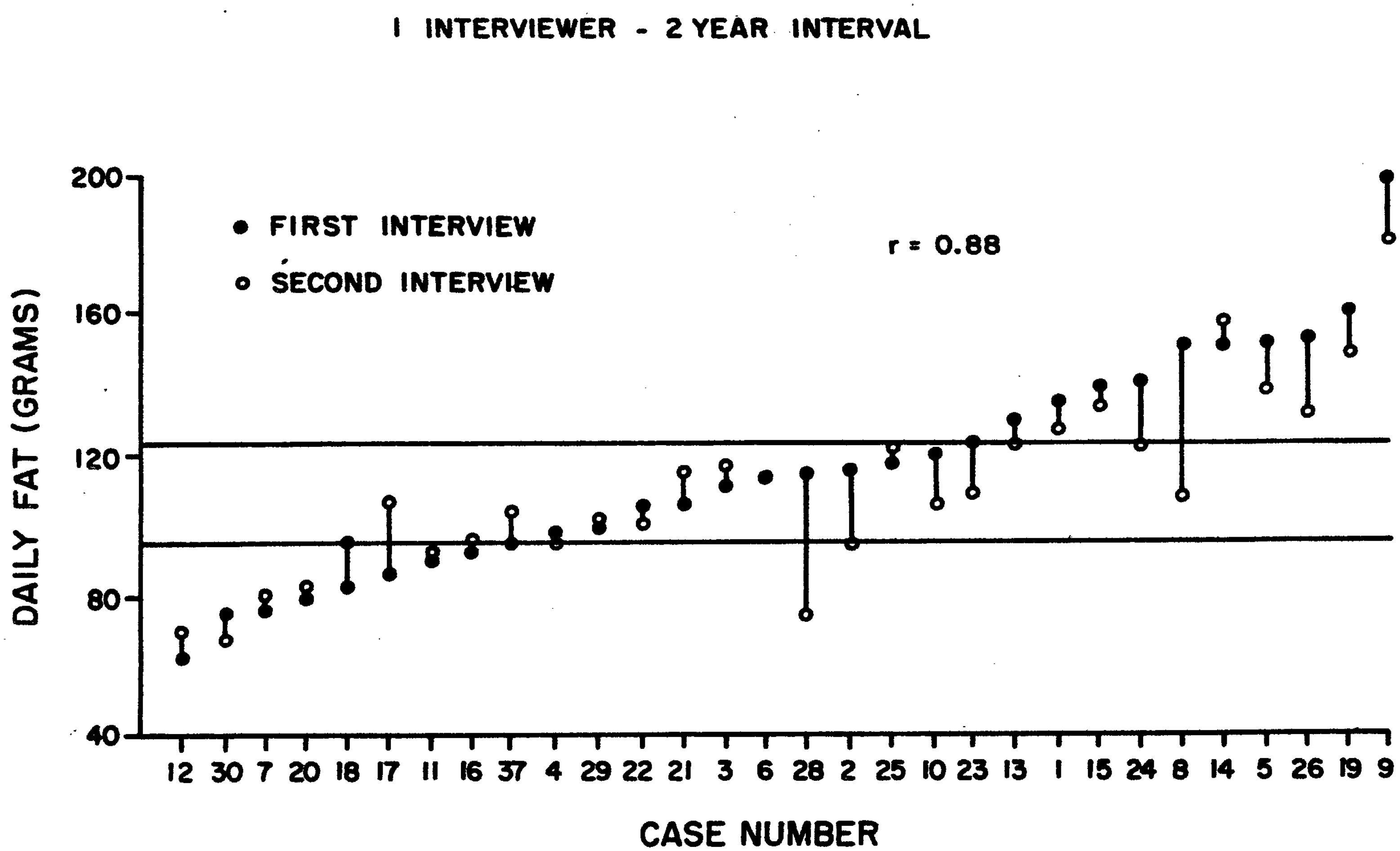
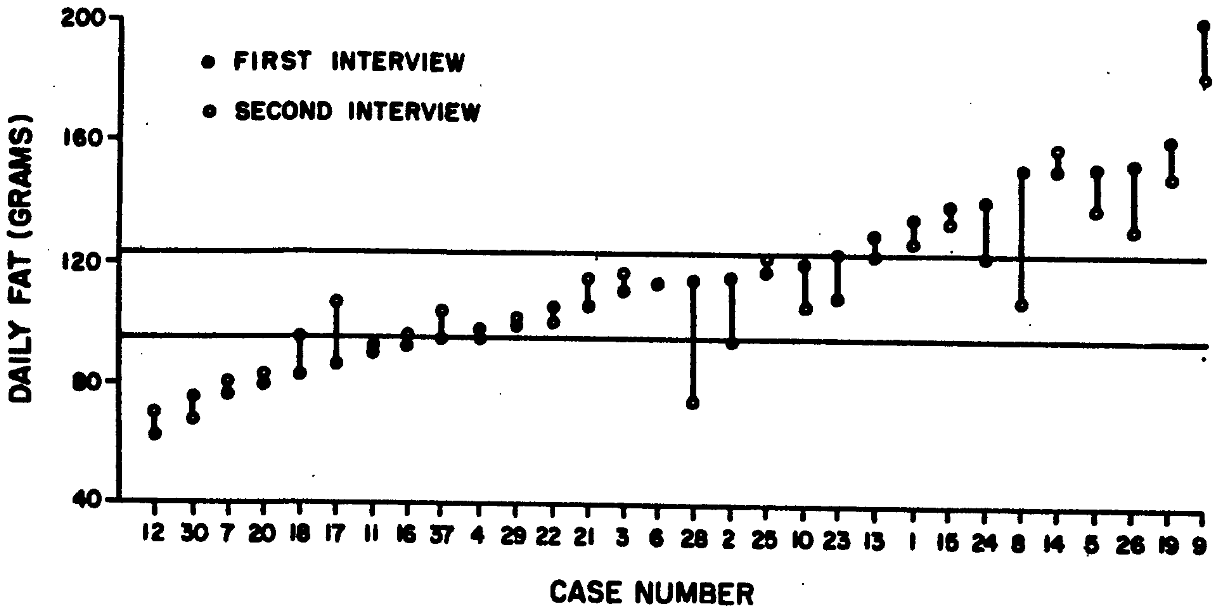

I INTERVIEWER - 2 YEAR INTERVAL
FIRST INTERVIEW
SECOND INTERVIEW
r = 0.88
DAILY FAT (GRAMS)
200
160
120
80
40
12 30 7 20 18 17 11 16 37 4 29 22 21 3 6 28 2 25 10 23 13 1 15 24 8 14 5 26 19 9
CASE NUMBER

Figure V

Figure VI


200
160
120
80
40
DAILY FAT (GRAMS)
● FIRST INTERVIEW
○ SECOND INTERVIEW
12 30 7 20 18 17 11 16 37 4 29 22 21 3 6 28 2 25 10 23 13 1 15 24 8 14 5 26 19 9
CASE NUMBER

BIBLIOGRAPHY

1. Enterline, D. E. and W. H. Stewart, Geographic Patterns in Deaths from Coronary Heart Disease, Publ. Health Rep. 71 849 (1958).

2. Lew, E. A., Some Implications of Mortality Statistics Relating to Coronary Artery Disease, Jour. Chron. Dis. 6 192 (1954).

3. Cornfield, J., Estimation of the Probability of Developing a Disease in the Presence of Competing Causes, Am. J. Publ. Health 57 601 (1957).

4. James, G., R. E. Dalton and A. S. Heslin, Accuracy of Cause-of-Death Statements on Death Certificates, Publ. Health Rep. 70 39 (1955).

5. Keys, A., The Diet and the Development of Coronary Heart Disease, Jour. Chron. Dis. 4 364 (1956).

6. Yerushalmy, J. and H. E. Hilleboe, Fat in the Diet and Mortality From Heart Disease. A Methodologic Note., New York J. Med. 57 2343 (1957).

7. Mann, G. V., G. Pearson, T. Gordon and T. R. Dawber, Diet and Cardiovascular Disease in the Framingham Study I. Measurement of Dietary Intake, Amer. Jour. Clin. Nutr. 11 200 (1962).

8. Dawber, T. R., G. Pearson, P. Anderson, G. V. Mann, W. B. Kannel, D. Shurtleff and P. McNamara, Dietary Assessment in the Epidemiologic Study of Coronary Heart Disease: The Framingham Study II. Reliability of Measurement, Amer. Jour. Clin. Nutr. 11 226 (1962).

9. Pearson, G., G. V. Mann and T. R. Dawber, Diet and Cardiovascular Disease in the Framingham Study III. Food Intakes, Amer. Jour. Clin. Nutr. (in press).

10. Marr, J. W., A Technique of Carrying Out Individual Weighed Diet Surveys, Proc. Nutr. Soc. 20 39 (1961).

11. Morris, J. N., J. W. Marr, J. A. Heady, G. L. Mills and T. R. E. Pilkington, Diet and Plasma Cholesterol in 99 Bank Men, Brit. Med Jour. i 571 (1963).

12. Rivin, A. U., J. Yoshino, M. Schickman and O. A. Schjeide, Serum Cholesterol Measurement: Hazards in Clinical Interpretations J. A. M. A. 166 2108 (1958).

13. Mann, G. V., A Method for Measurement of Cholesterol in Blood Serum, Clin. Chem. 7 275 (1961).

14. Watkin, D. M., E. Y. Lawry, G. V. Mann and M. Halperin, A Study of Serum Beta Lipoprotein and Total Cholesterol Variability and its Relation to Age and Serum Level in Adult Human Subjects, Jour. Clin. Invest. 33 875 (1954).

15. Page, I. H., The National Diet-Heart Disease Study, J. A. M. A. (in press).

16. Dayton, S., M. L. Pearce, S. Hashimoto, L. J. Fakler, E. Hiscock and W. J. Dixon, A Controlled Clinical Trial of a Diet High in Unsaturated Fat, New Engl. Jour. Med. 266 1017 (1962).

17. Rinzler, S., Nutrition in Relation to Heart Disease. The Anti-Coronary Club of New York., Arch. Environ. Health 5 60 (1962).

SOME CONSEQUENCES OF SOME ASSUMPTIONS WITH RESPECT TO THE PHYSICAL DECAY OF A CHAMBER AEROSOL CLOUD*

Theodore W. Horner*
Project Statistician, Booz-Allen Applied Research, Inc., 4815 Rugby Avenue, Bethesda, Maryland

The usual piece of equipment for studying the survival charactersitics of organisms suspended in an atmosphere is a gas-tight chamber controlled with respect to relative humidity and temperature. The mathematical formulation of the behavior of aerosol clouds injected into these chambers and the viability of organisms contained in the particles of these clouds are of great interest to aerobiologists. This paper is concerned with some of the consequences of a particular set of assumptions with respect to the physical decay of chamber aerosol clouds. In presenting the material, I will first touch on those aspects of chambers and aerosol clouds that must be taken into consideration in mathematical formulations. Biological recovery curves will be touched on next. A discussion of relationships among parameters associated with the physical recovery of the cloud will follow -- hitting first the mathematical characterization of the assumptions, then the mathematical relationships among the parameters and finally by means of slides, the relationships will be pointed up visually. The paper will conclude with a short discussion of possible applications of the work and an indication of further work that remains to be done.

Chambers vary enormously in size. Usually they are cylindrical in shape, being oriented either horizontally or vertically. Occasionally they may have a spherical or some other type of shape. The chamber may or may not be revolving. In using a chamber the procedure is to disseminate an aerosol cloud from a liquid slurry containing viable organisms into the chamber.

The aerosol cloud is composed of liquid droplets and the disseminating device produces a spray from the liquid slurry which is similar to the ordinary nose spray used to fight the common cold. Immediately after dissemination, the suspended particles start disappearing from the chamber due to gravitational fallout and impingement on the sides of the chamber. Since

---

*Report on work under Task 2 (Biological Aerosol Decay) or Contract No. DA-18-064-CML-2810 with the Program Coordination Office at Fort Detrick, Frederick, Maryland.

the larger particles fall more rapidly than the smaller particles, the distribution of particle sizes in the aerosol cloud changes with time. The usual assumption is that the particle number distribution is log normal immediately after dissemination. The fraction of those particles with radii between r and $r + dr$ is $f(r)\,dr$ where $f(r)$ is the frequency density function of the particle number distribution. Because of the differential fallout of the various sized particles, the particle number distribution does not remain log normal. The usual expression for differential fallout is

$$h(r, t) = \exp(-Kr^2 t)$$

where $h(r, t)$ is the fraction of those particles with radii between r and $r + dr$ that remain suspended at time t of those suspended initially. K is a constant that depends on chamber dimensions, gravitational acceleration and other factors. This formula which traces to Stokes law was first derived for stirred stationary chambers by Boyd*. Later Calder** showed that a similar formula held for revolving chambers.

Immediately after dissemination the aerosol particles undergo an equilibration process with respect to their moisture content and the chamber atmosphere. This process ordinarily is accomplished in about a second and so it is convenient to refer to <u>time zero</u> as that instant at which the equilibration process is completed. Both the equilibration and the dissemination process are quite drastic events in the life of an organism and so it is not surprising that many organisms which were viable in the slurry are dead at time zero. The organisms continue to die after time zero. The percentage of those organisms which were viable in the slurry, which remain viable at time zero is known as the <u>initial recovery percentage</u>.

The biological recovery percentage is the ratio, expressed in percentage form, of the number of viable suspended organisms at time t to the number of suspended organisms at time t. In this definition only the

---

*Boyd, Charles A., "The Theory of Sedimentation and Decay of Aerosols", Interim Report BLIR-7, Fort Detrick, July 1952.

**Calder, Kenneth L., "Some Theoretical Aspects of the Rotating Drum, Aerosol Chamber", BWL Technical Note 13, 1958.

organisms which were viable in the slurry are considered. We will represent the biological recovery percentage as B(t) and thus

$$B(t) = 100 \left\{ \frac{\text{Number of viable and suspended organisms at time } t}{\text{Number of suspended organisms at time } t} \right\}$$

Characteristics of the biological recovery curve B(t) as it varies with chamber size and shape, relative humidity, temperature, organism and slurry additives are of great interest to investigators studying the viability of organisms suspended in an atmosphere. The typical biological recovery curve when plotted versus time on semilog paper is concave upward. An estimate of the biological recovery percentage at time t depends on data from a sample of the aerosol cloud withdrawn from the chamber at time t.

There appear to be a number of empirical mathematical expressions that do an excellent job of fitting biological recovery data.. These expressions will often explain 99.5 per cent of data variability. The expressions generally have no theoretical basis and give rise to differing consequences. Thus inferences based on these empirical curves are always suspect.

As a step toward deriving biological recovery curves from a more fundamental foundation, BAARINC suggested some time back the heterogeneous initial recovery model. The model postulates that the concave upward curvature of semilog plots is due to nothing more complicated than:

(1) Distribution of particle sizes,

(2) Differential fallout of the various sized particles, and

(3) Higher initial recovery percentages for organisms contained in the larger particles.

These three assumptions are sufficient to generate the type of curvature normally observed. There appears to be no question about the first two postulates. The validity of the third remains to be proven, although it does appear to be quite reasonable to the aerobiologists with whom I have been in contact.

Essentially the heterogeneous initial recovery model states that the biological recovery percentage at time t is a weighted average of the biological recoveries associated with the various sized particles. The weights are the fractions of the suspended organisms that are contained in the various sized particles. These weights continuously change with time. The mathematical formula for the biological recovery percentage at time t is

$$(1) \qquad B(t) = \overline{B}(r, t) = \int_0^\infty \frac{Nf(r)\, h(r, t)\, ar^s}{\int_0^\infty Nf(r)\, h(r, t)\, ar^s\, dr}\, B(r, t)\, dr$$

where $B(r, t)$ is the biological recovery percentage for organisms contained in particles with radii between r and $r + dr$. The number of organisms contained in a particle of radius r is assumed to be proportional to the radius raised to the s^{th} power. Some work by Dr. William C. Day* at Fort Detrick tends to indicate that s may be different from 3.

The weights mentioned a few moments ago are indicated by the trapezoid drawn in the equation above. To point up the logic of these weights, we let N be the number of suspended particles at time zero. $Nf(r)\, dr$ is then the number of suspended particles with radii between r and $r + dr$ at time zero. Multiplication of $Nf(r)\, dr$ by $h(r, t)$ yields the number of suspended particles at time t with radii between r and $r + dr$. Further multiplication by ar^s yields the number of suspended organisms. Hence the denominator of (1) is the number of suspended organisms at time t and the trapezoid ratio is the fraction of suspended organisms contained in particles with radii between r and $r + dr$.

From equation (1), it is evident that characteristics of the biological recovery curve are intimately tied to the physical aspects of the cloud. In any case an understanding of these physical aspects must precede attempts to ascertain the validity of the heterogeneous initial recovery model.

*This work is described by Horner in a Biomathematics Analysis Note. Horner, Theodore W., Fort Detrick, Maryland, Biomathematics Analysis 5082, "A Relaticnship Between Spore Number and Particle Size", September 14, 1961.

What are reasonable assumptions and what are their consequences. It appears reasonable to assume that $f(r)$ is log normal and $h(r, t)$ is of the form $\exp(-Kr^2t)$. Further the mass of a particle, say $m(r)$ is probably proportional to the cube of the particle radius.

To check the validity of these assumptions and to estimate the relevant parameters is not easy for three reasons:

(1) The particle number distribution does not remain log normal. Part of the present investigation was designed to gain an understanding of the extent of this non-log normality.

(2) Chambers cannot be sampled at time zero and hence estimates of the parameters of the log normal distribution at time zero must be obtained by indirect means based on data collected after time zero.

(3) The physical recovery fraction of the cloud involves still another factor; namely, the mass of the particle.

The physical recovery fraction, normalized to 100 per cent recovery at time zero, is given by the formula

$$R(t) = \frac{\int_0^\infty Nf(r)\, h(r, t)\, br^3\, dr}{\int_0^\infty Nf(r)\, br^3\, dr} \tag{2}$$

The mass of a particle is $\overline{m}(r) = br^3$. The total mass suspended at time t and time zero respectively is given by the numerator and denominator of (2).

Following the assumptions made earlier, the recovery fraction is a function of the mean (u), the variance σ^2 of the initial particle number distribution and the chamber constant K. Thus

$$R(t) = R(t;\ u,\ \sigma,\ K).$$

In gaining information about the physical recovery curve one can, on the basis of aerosol samples, do several things. Thus you can:

(a) Estimate the physical recovery fraction at time t.

(b) Estimate characteristics of the particle number distribution such as the mean and the variance of $Y = \ln r$. At time zero, y would be a normally distributed variable.

Estimates of $R(t|u, \sigma, K)$, $E(y|u, \sigma, K, t)$ and $Var(y|u, \sigma, K, t)$ can be obtained from the data of aerosol samples, where u and σ are parameters of the log normal distribution at time zero. Knowing these three quantities, it would be desirable to know u, σ, and K. This would provide a basis for checking the validity of the theory concerning K and the log normal, $\exp(Kr^2t)$ system in general.

Our work has led to several equations that should be useful in this connection. As mentioned earlier, y was defined as $y = \ln r$. The variables v and q will now be defined as

$$v = (1/\sigma)(y - u)$$

and

(3) $$q = -(1/2) \ln Kt - u.$$

Using these definitions, the physical recovery fraction at time t can be written as

$$R(t \mid u, \sigma, K) = R(q, \sigma)$$

where

(4) $$R(q, \sigma) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp\left[-(1/2)\, v^2 - e^{-2(q-3\sigma^2 - v\sigma)}\right] dv$$

Thus initially, the tabulation of R would have required the four quantities t, u, σ, and K to be taken into account. The q formula relates t, K and u and thus tabulation becomes simpler in that R needs to be computed only as a function of q and σ. One of the slides a little later will show the relationships among R, q, and σ^2. An approximation formula

has also been developed for computing q as a function of the physical recovery percentage and σ^2. This formula is useful in generating a starting value of q for iteration procedures for the solution of q given R and σ. The approximation formula is

$$(5) \qquad q = -(1/2)\ \ell n(-\ell nR) + \left\{A + B\ \ell n\left[-\ \ell n(1 - R)\right]\right\} \sigma^2$$

where A and B are appropriate constants.

Using this formula suppose q is calculated given R and σ^2. If the approximation q is now used to calculate R using equation (5), the calculated R will not differ from the original R by more than 0.02 for the range of σ values pertinent to the present investigation.

Some additional equations are listed below.

$$(6) \qquad E(y \mid u, \sigma, K, t) = u + \sigma E(v \mid q, \sigma)$$

$$(7) \qquad Var(y \mid u, \sigma, K, t) = \sigma^2\ Var(v \mid q, \sigma)$$

$$(8) \qquad R(t, u, t, K) = R(q, \sigma)$$

where the frequency density of v is

$$m(v) = \frac{(1/\sqrt{2\pi})\ \exp\left[-(1/2)\ v^2 - e^{-2(q - v\sigma)}\right] dv}{(1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp\left[-(1/2)\ v^2 - e^{-2(q - v\sigma)}\right] dv}$$

Let us look at equations (7) and (8). These are two simultaneous equations. Having estimates of Var y and R(t) available at time t, one can solve for estimates of q and σ. Estimates of q and σ make possible an estimate of $\sigma E(v \mid q, \sigma)$. This latter estimate, when coupled with an estimate of

$E(y \mid u, \sigma, K, t)$, makes possible an estimate of u. Equation (3) can be solved for the chamber constant K. This constant is a function of q and u. Since estimates of q and u are available, it is now possible to estimate K. Thus from estimates of q and u are available, it is now possible to estimate K. Thus, from estimates of R, E_y and Var y at time t, one can determine u, σ, and K.

The relationships among these quantities can best be pointed up by means of graphs. For the graphs to be meaningful, it is necessary to make appropriate choices for the parameters u and σ o the log normal distribution of particle sizes at time zero. For u, the range from 0.5 to 5μ seems appropriate in the subject matter area to which this investigation is related. Similarly, an upper bound for σ appears to be in the neighborhood of 1.5. To arrive at this latter number we define two radii r_1 and r_2 such that 50 and 84.13 per cent of the particles at time zero have radii less than r_1 and r_2, respectively. Under these assumptions,

$$\ln r_2 = u + \sigma$$

and

$$\ln r_1 = u$$

Hence

$$\sigma = \ln (r_2/r_1)$$

and

$$r_2/r_1 = e^{\sigma}$$

When σ is 1.5, the ratio of the two radii is 4.4817. Thus the radius associated with the 84.13 per cent point is 4.48 times the radius for the 50 per cent point when $\sigma = 1.5$. Thus $\sigma = 1.5$ appears to be a reasonable upper bound for the subject matter under investigation. In developing tables, the values of σ shown below were used.

σ	r_2/r_1
0.40547	1.0
0.69315	2.0
1.09861	3.0
1.38629	4.0

Figures 1, 2, 3, and 4 [figures are at the end of this article] illustrate various parameter relationships. Figure 1 shows the relationships among the physical recovery fraction, q and σ^2. Each line is associated with a different physical recovery fraction. These lines are almost but not quite straight; the curvature is most pronounced in the lines associated with the lower physical recoveries.

Figure 2 shows the relationships between q, R, and Var y, each line being associated with a different physical recovery fraction. Again these lines are almost straight; the greatest curvature being associated with the lowest physical recovery fractions.

The third figure shows the relationships among (1/Var v), Var y and R. This graph can be used to estimate σ^2 from estimates of R and Var y. The estimate of σ^2 is the product of the estimates of Var y and (1/Var v). Thus suppose R at time t is estimated as 50 per cent and Var y is estimated as 0.43. The value of (1/Var v) is then estimated as 1.12. The estimate of σ^2 is then (0.43) x (1.12) = 0.48.

The final figure shows relationships among ($-\sigma$ Ev), R and the standard deviation of y. The estimate of u is found by adding ($-\sigma$ Ev) to the estimate of Ey. Again suppose R = 0.50 and $\sigma_y^2 = 0.43$. In this case $\sigma_y = 0.66$ and the estimate of ($-\sigma$ Ev) is 0.065μ. Thus the estimate of u is found by adding 0.065 to the estimate of u_y.

Hopefully, the relationships which have been developed will lead to

(1) Quick and efficient estimation procedures for the parameters which characterize physical decay in aerosol chamber trials,

(2) Aids useful in designing and interpreting chamber experiments

(3) Procedures for testing the validity of the common assumptions with respect to physical decay, and

(4) Ways of evaluating the bias in methods of estimating biological recovery percentages which employ mass tracer data.

Finally and most important of all, these relationships constitute a start toward developing methods for testing the validity of the heterogeneous initial recovery model for biological recovery.

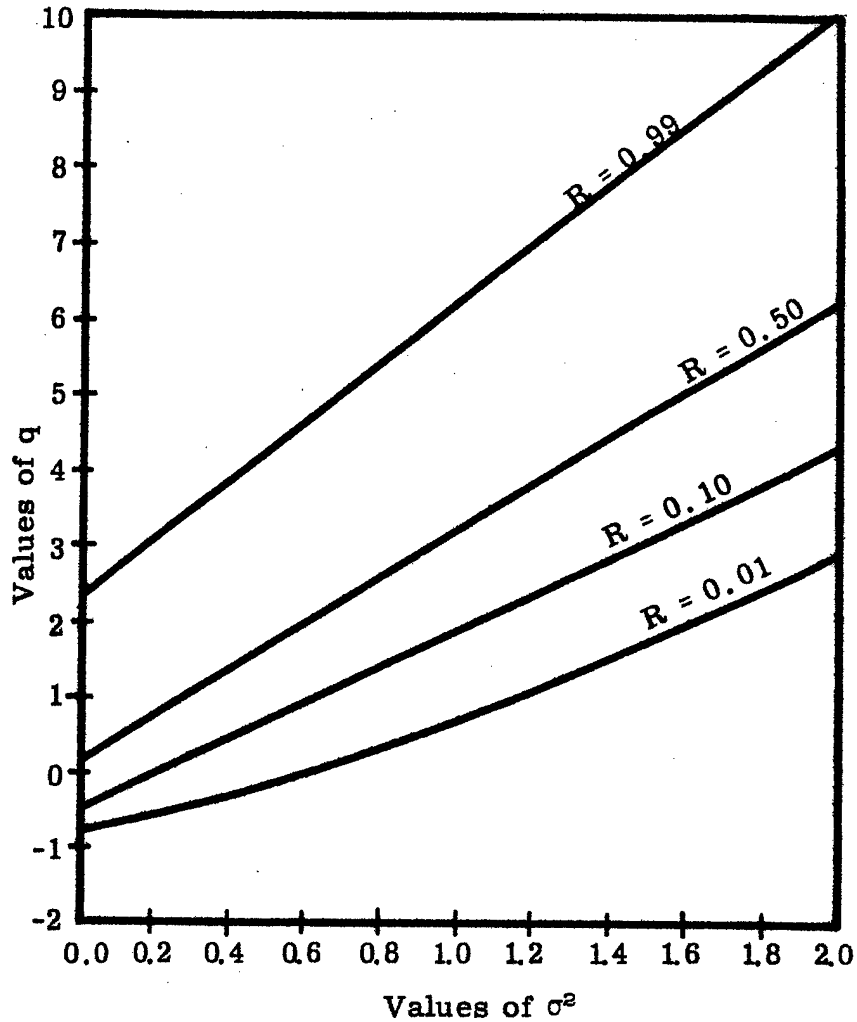


Figure 1. Graph of q versus σ^2 for various values of R where

$$R = (1/\sqrt{2\Pi}) \int_{-\infty}^{\infty} \exp\left[-1/2)\, v^2 - e^{-2(q-3\sigma^2-v\sigma)}\right] dv.$$

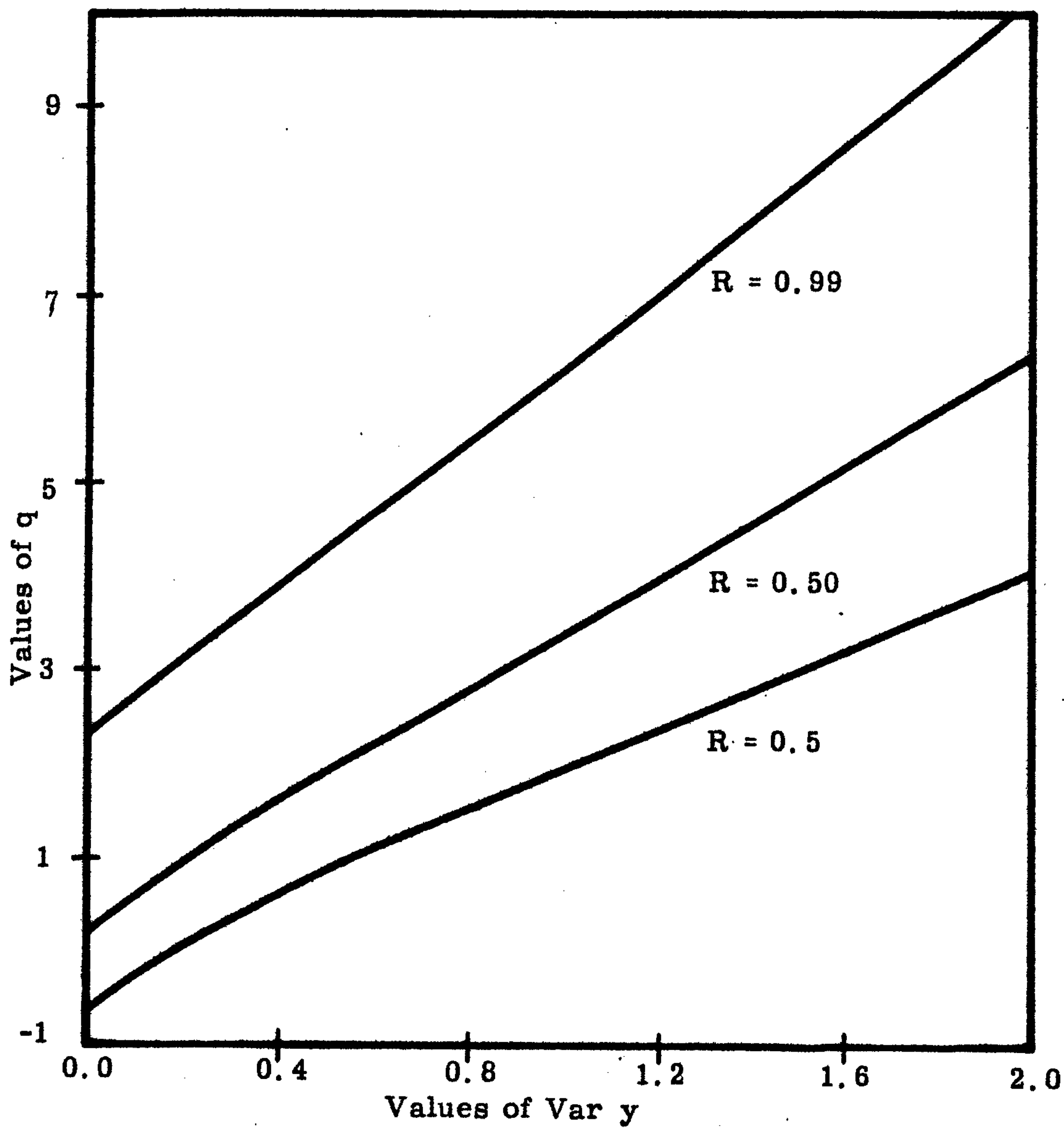


Figure 2. Values of q versus Var y at various physical recovery fractions.

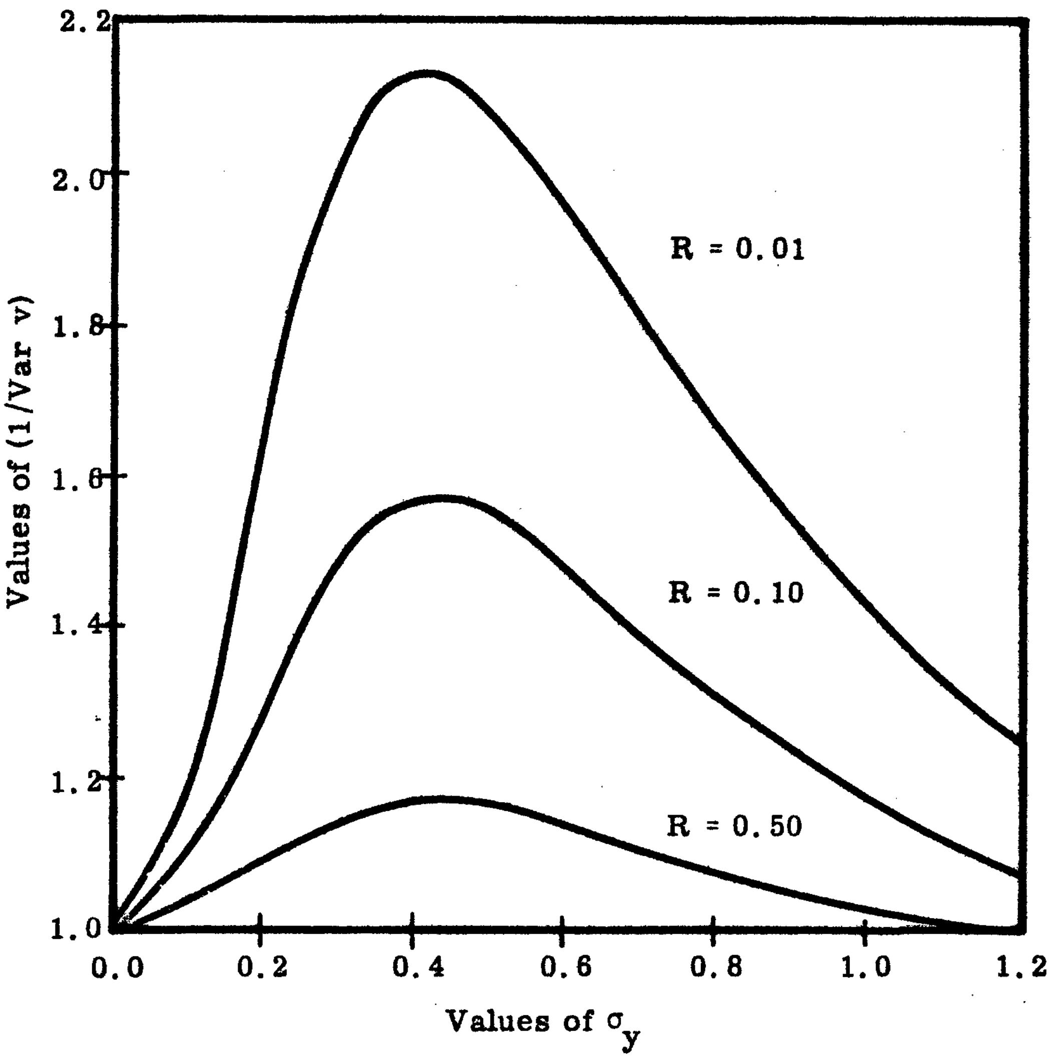


Figure 3. Values of (1/Var v) versus σ_y at various physical recovery fractions.

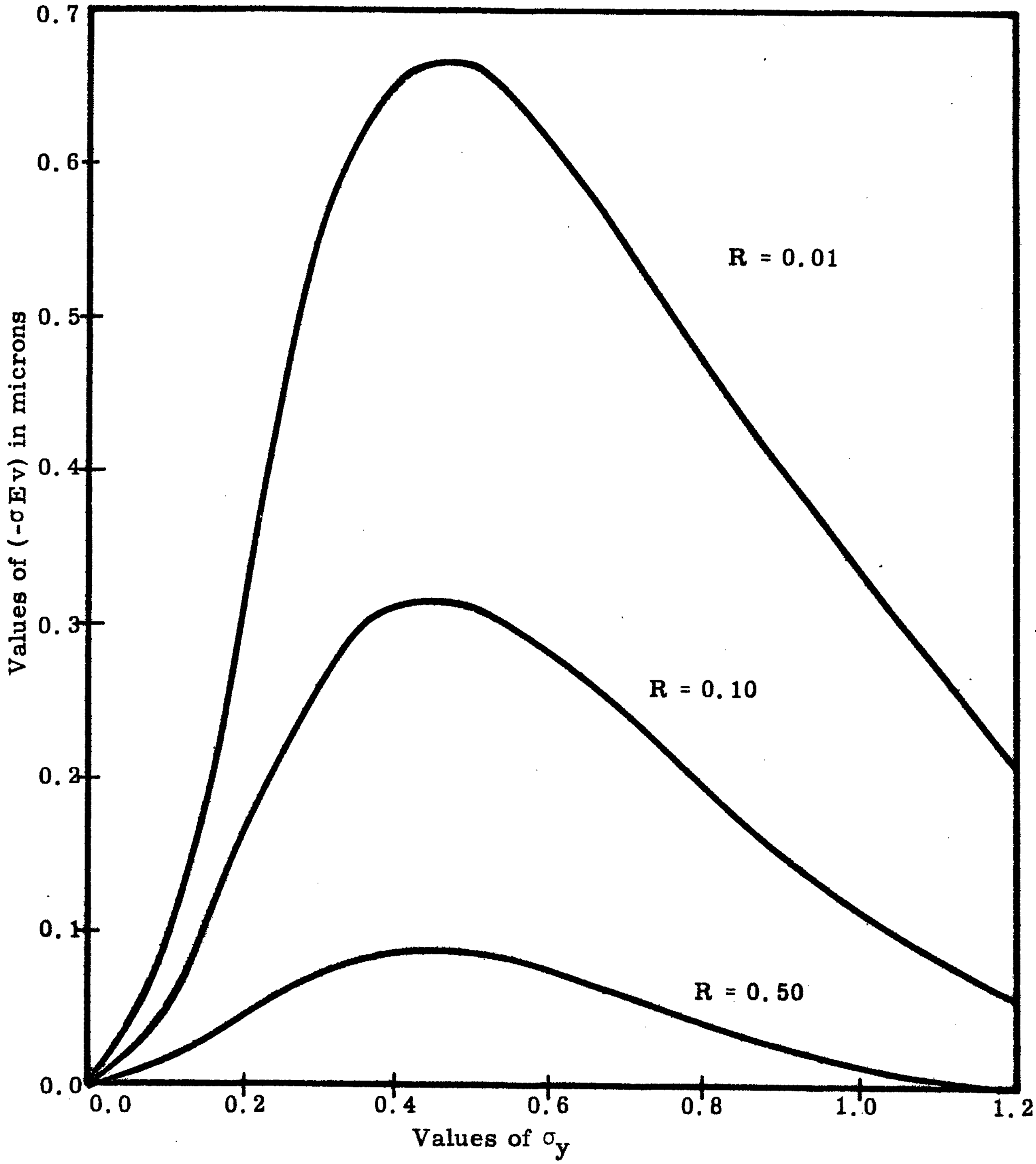


Figure 4. Plots of $(-\sigma E v)$ versus σ_y at various physical recovery fractions.

THE ROLE OF INTUITION IN THE SCIENTIFIC METHOD

Nicholas M. Smith
Research Analysis Corporation

ATSTRACT. By virtue of Huygens' careful appraisal of the Newtonian method of inquiry, a logical fallacy has long been detectable in Newton's contention that he had deduced truth from observations of nature. The logical fallacy is concerned with two aspects over which empiricism has no control: (1) the observation that what constitutes a fact in the test of a theory is not determined by empirical principles alone, and (2) the observation that the constructs of science cannot be demonstrated to be other than sufficient with respect to the so-called facts. The necessity--that is, the uniqueness--of these constructs can never be established. The realization has gradually emerged that whatever is "true," "valid," or "warrantable," is not to be determined by absolutely singular discoveries produced in flashes of insight, but by selection among alternative creative insights on the basis of systematic tests constituting the reasoning process. The possibility of alternative insights of equal predictive applicability makes necessary the imposition of some principles other than empirical ones to decide among them. These principles are intuitive and categorical. That is, there exists a set of cognitive controls (of which empirical tests are members) which are established for the sole purpose of preventing ambiguity. Some of these principles have appeared in modern and contemporary science, notably the principle of relativistic invariance, which can be traced directly to the need of preventing procedural ambiguity introduced by transformations. There are other important introspective controls which delimit forms acceptable for application in the cognitive act. Many of these introspective principles are to be found separately in contemporary theories of value. We have formulated a theory of these cognitive controls based on our attempts during the past two years to establish the foundations of a rational methodology for systematic and organized prescriptive activities, that is, the decision process in all its generality. Science, as a decision system which has as its purpose the production of predictive theories, is shown to be a reduction of the more general axiologies. Consequently, many of the important so-called laws of science are not singular discoveries of properties of the external world entirely, but are in addition necessary properties of admissible forms which may serve as objectifications for applicable symbolic scientific models. Among these principles may be some of the most cherished scientific disccveries: the relativistic properties of space-time, the conservation of momentum, the conservation of energy, the second law of thermodynamics, and the Heisingberg uncertainty principle.

INTRODUCTION. The purpose of this discussion is to present the findings of a study* conducted at the Research Analysis Corporation for the past several years concerning the nature of the rational or cognitive process. This study has revealed that intuition (or introspection, as we shall call it) plays a much greater role in the process of rational thinking than we had heretofore suspected. The nature and complexity of the subject is such that a detailed presentation in a systematic and convincing step-by-step procedure would require many hours. We shall therefore resort to a presentation of our material in the time available in the form of an elaborate abstract. It cannot be hoped that such a shortened version of our presentation can be wholly convincing. It is hoped, however, that the attention of the reader will have been directed to some shortcomings in prevalent notions of the scientific method as applied to the design of experiments. We also hope that our method of resolving these problems will seem plausible and that your own interest in this exciting field will be aroused.

The study I am describing has been motivated by a search for the foundations of management science. In the term "management science" I mean to include such other terms as operations research, operations analysis, industrial engineering, economics, and the like. Those who practice these professions are not in complete agreement as to a statement of their mission; but in general these ingredients will be found in any definition: management science is somehow to provide a client with aids--quantitative or otherwise--to one of his decision processes. Or, the analyst may even go so far as to recommend specific decisions to the client. These aids, or these recommendations, are formulated with respect to the client's value system; it is further claimed, either implicitly or explicitly, that the management scientist employs a method which will somehow lead to better decisions. These definitions are charged with highly significant but poorly defined words. These words are "decision," "values," "method" and "better." The search for an understanding of the ideas behind these words has triggered an escalade of theoretical projects.

In the first place we have committed our interest to the field of practical decisions and therefore have become interested in the theory of decision algorithms. To many persons who practice our profession this subject may appear as the sole content of management science. This general field covers such divisions as mathematical programming, queueing theory, logistics theory, game theory, etc. The central commodity in terms of which decision

* This paper describes work done under RAC Study 5.4, "Valuation and the Cognitive Process," by N. M. Smith and M. C. Marney.

makers operate in reaching their decisions, however, is value. Since values are the determinants of decisions, a whole new theoretical field in the theory of value is developing. Value theory, on the other hand has drawn attention to the decision process in context of system. One cannot understand the act of evaluation without understanding the nature of the system in which the evaluation process is undertaken. This situation thus leads to a third theoretical project--the theory of selective systems.

Finally one must turn to the question of the validation or warranting of values, policies, and ethical systems. The question of such warrant, together with questions concerning the adequacy of the methodology of professional management scientists, have drawn out interest into the general theory of cognitive processes.

RECONSTRUCTION OF PHILOSOPHICAL FOUNDATIONS. The first impediment one encounters in the search for a method of warranting a value-decisions process (or, as we shall call it, a prescriptive process) is the conculsion that contemporary scientific method is inadequate. This inadequacy arises because one of the chief controls in the scientific method is a predictive process. One attempts to test or "warrant" a scientific theory by means of predicting future observations. A comparison of actual observations, with a suitable definition and range of measurement, will then define a warrant for a scientific theory. In the prescriptive process, on the other hand, one cannot confirm the adequacy of a value or policy by predicting one's own decisions, since these values are the indices which determine these decisions. Such a test would merely demonstrate a degree of consistency with respect to policy. We have gradually become aware that the prescriptive process is somehow different from the predictive process. In subsequent developments of the theory we have found that the differences and similarities between prescription and prediction are fairly complex, as I shall attempt to demonstrate.

Failing to find a rational prototype for the validation of prescriptive processes, we turned to a survey of historical and contemporary ethical and value theories. Although we found literally dozens of philosophical schools which purported to provide a means of selection and control of ethical systems, all of them exhibited inadequacies of various kinds. Failures of these systems and the historical failures of older ethical systems and scientific methods have occurred in a characteristic pattern: ultimately they have been confronted with situations which could not be resolved by the principles espoused.

It may also be observed that, to a large degree, it has been supposed that three great rational methodologies are treated as if they were separate processes. I am referring to (1) axiomatics, a selective system that produces valid formal systems, (2) scientific method, a selective system that produces predictive theories, and (3) axiological method, a method that produces prescriptives, policies, ethics. It has been assumed that axiomatics may be adequately controlled entirely by the rules of logic. On the other hand, the history of the scientific method has been characterized by the accretion of both logical and empirical controls and, in modern science to some degree, by the injection of intuitional controls. The axiologies have been presumed to have been controlled entirely by intuition. There have been, of course, attempts to approach ethics and values from a naturalistic viewpoint as predictive entities, but these studies can be shown to be concerned with value systems as objects, wherein our chief concern has been with value systems as subjects. (That is, what should my policy and my values be in order to determine my decisions?)

Having failed, then, to fine a rational prototype for the warranting of the prescriptive process, we have been forced to attempt a reconstruction in the philosophical foundations of the rational method in order to incorporate axiologies into the group of systematic rational pursuits. This reconstruction has taken the nature of a synthesis among modern scientific methods and contemporary and historic value and ethical theories. It promises, besides its direct application to axiology, to yield additional enlightenment on axiomatics (that is, the control of mathematical method) and the scientific method. This intimation is pertinent to the objectives of this conference and represents the specific subject of my discussion.

THE DEVELOPMENT OF A METATHEORY. It is desirable to distinquish between metatheory and object-theory. An object-theory is a theory which objectifies, or externalizes, objects. Such theories create the following types of objects: the objects of mathematics and logic are sets of self-consistent formal statements together with their consequent theorems; the objects of science are particular predictive theories and the elements thereof; and the objects of axiologies are particular policies that determine or prescribe practical decision. A metatheory, on the other hand, is a theory about object-theories. In particular it is a theory about the methods of admission or control and warrant of object theories. Our theory is a metatheory in which we are attempting to synthesize the metatheories of mathematics, science and axiology under one conformal perspective.

Any such theory presupposes, explicitly or implicitly, certain ontological and epistemological commitments, i.e., commitments as to what constitutes existence and knowledge respectively. Central among our commitments is the notion of relativism in three facets: the first is onotological relativism. This refers to the doctrine that existence of an object-construct is determined by its testability in principle or its connectability by inference to other object-constructs which are testable in principle. In other words, one rejects the notion of things-in-themselves or concepts which, by their very nature, are not subject to test. The term "test," of course, refers not only to empirical tests but to intuitive and formal tests as well.

The second facet is relativism in epistemology. This refers to the doctrine that certainty of knowledge of object-constructs, i.e., the establishment of apodictic truth (truth by necessity) is not obtainable. One must observe that the proofs of validity or warrantability of any scientific theory merely determine the efficiency of that theory in coordinating and clarifying the information obtained under a consistent predictive format. There is never any certainty that some other theory may not be developed which would describe the observations equally well or better; nor is there any certainty that the presently accepted theory will be adequate with respect to any future information that may be obtained. The consequence of these observations is that absoluteness at the object level is not meaningful.

A third facet is perspective relativism. This refers to the doctrine that an absolute reference for the judgment of object-theories is not obtainable. As we shall see in a moment, the consequence of this commitment is Einstein's principle of invariant transformations.

A second commitment presupposed by our meta-theory is that the sole function of the metacognitive process is the assurance of decidability of object-statements--that is, decidability with respect to their admissibility. The concept here is that relativism in object-space leads to degrees of freedom. Decision as to which object-constructs in this range of freedom are to be admitted must be accomplished in terms of some metaprinciple or control. This control then becomes a new absolute replacing the absolutes relinquished at the object level. That is, the controls are categorical, and they are metacontrols. The consequence is the conclusion that <u>ambiguity is the sole motivation for decision</u>.

<u>THEORY OF COGNITIVE CONTROLS</u>. There are, however, many kinds of ambiguities and each type of ambiguity necessitates a corresponding control. As we have said before, these controls

are categorical and their sole function is to resolve ambiguities of the class to which they apply. We have classified the controls in terms of three factors which we call formal, extrospective, and introspective. Besides these reflexive or internal controls there are also sets of external controls which we refer to as evolutionary and aesthetic. One of the great difficulties in developing an acceptable metatheory is collection of provision for selection among alternative object-theories which purport to apply to the same problematic situations. This selection is accomplished by means of evolutionary control--a generalization of the Darwinian principle--and aesthetic control (elegance). The evolutionary controls (fruitfulness, adaptability, and survival) represent ultimate commitments. Since the general thesis of this presentation can be developed without an elaboration of these important concepts, and since time does not permit such an elaboration, we shall forego any further discussion on these topics.

Central to our theory is the concept "objectification." Objectification represents the emergent result of a creative act which externalizes, at the level of a cognitive agent or self, a set of new conceptual entities or object-constructs on a trial basis as an act of policy and subject to a warrant to be established for predictive or prescriptive purposes by a set of cognitive controls. In this viewpoint all rational process is undertaken in terms of object-constructs, a special class of object-constructs being theories or models.

The formal controls of an object-construct apply to its format or formal properties. They insure admissibility under tests of consistence, completeness and independence.

Extrospective Controls. There are two acts in the extrospective control. One is the determination of the criteria of fact--that is, the selection of the specification of what constitutes a relevant fact based on a formal objectification selected among an indefinite set of objectifications as an act of policy. The criteria of fact becomes a filter through which extrospection is admitted as relevant to the problematic situation at hand. Thus, in the act of its admission, any "fact" has formal, introspective and extrospective components. There is no such thing as a purely extrospective fact. This supports the views of contemporary philosophers of science. So let me repeat: this conference, concerned as it is with the design of experiment, or as I have called it, the criteria of fact, is concerned with much more than extrospective information or data. In particular, it is concerned with formal and introspective (that is, intuitive) properties. Now "extrospection" means a looking outwards, or receptivity to

information processed through transducers and subsystems whose outputs are presented to mediation at conscious level. On the basis of the objectification or model one undertakes a prediction, that is, a symbolic projection forward in time beginning with an extrospectively determined initial state and in terms of a specific objectification. This leads to an expectation. A significant discrepancy at a later time between expectation and extrospection engenders extrospective ambiguity. In order to define extrospective ambiguity, one must first select (a) a range of initial admissible expectations, (b) a range of admissible divergencies between expectation and extrospection at a later time, and (c) a frequency measure. We can now define extrospective ambiguity as follows: a set of final expectations and extrospections are empirically nonambiguous if, and only if, a set of histories all beginning with initial states in the admissible range are examined and are found to contain a subset of final states lying in the admissible range around expectation, such that the ratio of the number of final admissible histories to the number of initial admissible histories is equal to or greater than the frequency measure.

The decisions as to the admissible initial and final ranges and the frequency measures are determined by the problematic situations which are desired to be resolved by the objectification. This range of application represents an aesthetic decision. One could, for example (see Table 1), set the frequency measure equal to zero, in which case he is saying he is indifferent to the correspondence between expectation and extrospection. He then becomes, by this aesthetic orientation, primarily concerned with the formal properties of his objectification. That is, he becomes a mathematician. He maintains an interest in the residual substantive properties of his constructs as exhibited by his attention to the nature and efficiency of his notation.

If the range of problematic situations desired to be faced includes prediction of situations, then the frequency measure is set at a non-zero value. We shall call such a person a scientist provided that he has also set his norms with respect to action implied by his objectification at null values, such that he is indifferent to such action. If he becomes aesthetically oriented completely toward action with respect to all immediate and mediate problematic situations, he will, in general, find that he has greater difficulty in satisfying all of the cognitive controls and hence, facing more restrictive constraints, must reduce the scope of comprehensiveness of his models. A primary control is that of practicability. What is practicable with respect to an action problem may be oversimplified with respect to a predictive problem. What is practicable to a scientist may be impractical

Table 1

A UNIFIED META-CONTROL SYSTEM

Metasystem	Operation	Range of problematic situations	Aesthetic Decisions	Scope of objectifications practicable
Axiology	Retrodiction	All practical problems	All norms effective	Most severely restrained, most reduced
Science	Prediction	Specific predictive situations	Action norms at indifference	Restrictions moderate, richer range of objectifications
Axiomatics	Formal extension	Consistent axiomatic systems	Action norms at indifference, extrospective ambiguity measure at null	Least restricted, richest in formal content

to a man of action, etc., the objectifications becoming correspondingly richer as one goes from axiology, to science, to mathematics, as the cognitive controls become, in some sense, degenerate.

There is also a very important difference between the viewpoint of prescription and the viewpoint of prediction--that is, the prescriptive operation, although it may warrant its objectification or model predictively, when it is used in prescription it is turned around and used retrodictively. Now, retrodiction is not the exact reverse of the predictive process. It is this difference between retrodiction and prediction which makes science and axiology acquire complementary characteristics. One is said to be adjoint to the other.

This property has very important philosophical as well as methodological implications. In particular, the primal or predictive viewpoint represents the view of a construct as an object whereas the complementary or dual can be interpreted as a representation of the construct as a subject. Thus, in terms of value theory, predictive value theory is a system by means of which one observes the decisions of another person as data and makes a theory the value system of that person as an object. On the prescriptive side of value theory, one is concerned with one's own values as determinants of one's own decisions. It is this process that is retrodictive.

Introspective Controls. Time will not permit a detailed discussion of introspective controls. We shall endeavor, however, to say enough about these so that their function and importance can be realized. Let us look at perspective control. This is the direct application of our epistomological commitment to perspective relativism. One may refer a statement in an objectification to a particular context of coordinate systems. Ultimately, they may be transformed into another and a description made in terms of another coordinate system. If this transformation depends upon the procedure or path taken from one system to another, one would naturally get a different result from the transformation depending upon the path taken. This would result in what we may call perspective ambiguity. If there existed an absolute point of reference, then a natural algorithm for transformations would be indicated. One would simply transform from the first coordinate system to the absolute origin and from there to the new coordinate system. In the absence of any such absolute perspective, one must limit the transformations to those having a particular property.

We seek a class of transformations which do not lead to ambiguity, regardless of the procedure or path taken. These are

called invariant transformations and they result in a formal description in the new coordinate system which is identical to the formal description in the old coordinate system. This is the principle of invariance. While it has a rather abstract title, and while the discovery of invariant transformations may sometimes be difficult, the intent and meaning of the principle is very simple. It says merely that one must avoid procedurally induced ambiguities.

In a space-time transformation of a physical theory, this leads directly to the Lorentz-Einstein conditions for a space-time transformation. Now it is also true and also of interest that if one looks at an object-space determined by a Markov stochastic system and asks for a nonambiguous or invariant transformation of a velocity in a Markov space (i.e., the velocity of movement of a probability configuration), half of the conditions for an invariant transformation emerge as a result. The adoption of the second half, as necessary for an invariant transformation, is equivalent to the introduction of the set of imaginary probabilities which, together with the real Markov probabilities, are to be associated with each transition. The results*, which may not surprise you by now, are none other than, again, the Lorentz-Einstein transformation equations in the space defined by the Markovian system.

Before the time of Einstein, science and axiology were presumed to be entirely separate, science being the province of empiricism and formal logic, whereas axiology, separate and disconnected, was the province of intuition. Then Einstein shook the very foundations of physical theory by a brillant and successful modification of the cherished concepts of space and time--a modification which depended not on empirical discovery but upon application of an intuitional requirement.

Even today the commitment to empiricism is sufficiently strong, and naive realism is so firmly established, that the full significance of Einstein's principle is not realized. This principle does not refer to a singular discovery of a property of the external world, but instead to a necessary property of admissible forms which may serve as objectifications for applicable symbolizations. We are constrained to think in terms of perspectively invariant forms, or we are inevitably led to ambiguity. Einstein, having achieved a nonambiguous formulation of mechanics, was then able to proceed to show a relation between energy and mass. The relation between energy and mass is not a substantive consequence of relativistic invariance; it is merely a formal result educed by an enlightened procedure which was made possible by a form of nonambiguous thinking.

*Smith, Nicholas M., "A Calculus for Ethics: A Theory of the Structure of Value," Behavioral Science, Vol. 1, Nos. 2, 3. 1956.

Other Implications of Invariance. Einstein's invariance has other far-reaching implications, particularly when we generalize the principle to state that all formal objectifications must be invariance with respect to significant transformations. "Significant" transformations are those in which the ambiguity arising from noninvariancy will be distinguishable from the range of admissible extrospective error. Generalized invariance has particular importance and implication in value-decision theory. One demands by application of this principle that the transformation of decision from a present to a future state by means of the Chapman-Kolmogorov transformation shall lead to a form of the value-decision equation identical with the initial one. If this were not true, then the decision indicated by the value-decision equation would depend upon the procedure in which a decision was staged into parts for analysis. The requirement of invariance with respect to time-translation transformation is insured first by the nature of the Chapman-Kolmogorov equation, and second by the placing of an important restriction on the decision operator. This restriction is one of commutation. A decision operator which commutes through the stages of decision process will permit an invariant transformation of the equation as applied from one point of reference in time to another. This property is also known by another name. It is the principle of optimality of dynamic programming. The latter is connectable to Euler's Weirstrasse and Legendre conditions of steepest descent algorithms.

It may also be shown that the Chapman-Kolmogorov equation, as it enters into value theory, introduces a concept analogous to momentum by virtue of the fact that the value equation is analogous to the conservation of momentum. Again the selection of a model in which the Chapman-Kolmogorov equation applies has been based upon the need for an invariant model as a starting point for the building of a theory. It also may lead, one adds, to a suggested generalization or modification of the law of conservation of momentum.

Other Introspective Controls. There are other introspective controls, each in its way fully as important as the principle of invariance; and each, when stripped of technical verbiage, merely assures nonambiguity and therefore decidability in the object-model.

One of these controls refers to the context of an object-construct. It requires that the context be specified in order to complete the meaning of the construct and it further specifies that an object-construct may have only one context, since if it had more than one context, it would be ambiguous. This particular

control, a modification of the Russell-Whitehead theory of types, can be expected to have important significance in the removal of certain kinds of paradoxes from modern logic.

Another introspective control, which is a direct statement of onotological relativity, constrains all object-constructs to those which are testable in principle. A third control requires furthermore that the test of the construct must not only be attainable in principle, it must be attainable and interpretable in terms of finite processes. This control will rule out infinite processes and continuous time-space as directly applicable to substantive constructs. Such concepts must assume a secondary status--that of operating constructs which guide the interpretation of finite extrospection in the context of a particular objectification. Examples of such secondary or operating constructs are: the wave functions of wave mechanics (which operate away in the act of evaluating a measurable entity), the concept "true" probability, which is never attainable; the optimum in a decision process, which is never achievable; also included is the class of decision variables as contrasted with the class of object variables.

The effect of introspective controls is to restrain the selection of object-models which are admissible for serving as the formal content of object-constructs. It therefore should come as no surprise that the form of all successful theories (that is, theories which prove to be admissible under extrospective, introspective and formal tests) will shown strong analogies.

Nor is it surprising that scientists have discovered introspective principles in the course of empirical investigations and have believed them to be part of the extrospective content of their observations.

This is not to say that these explicit principles, when they appear, are wholly intuitive, but rather that they are the consequents of intuitive requirements. A successful theory--no matter in what terminology it is formulated--will contain these principles in order to be nonambiguous.

Modern mathematicians have rediscovered Einstein's principle of invariance recently and have given it a name implying an extrospective connotation--they call it the principle of causality, and further, go so far as to say that it is the basic principle of classical physics.

These sets of controls alone are not sufficient to determine decidability. They are reflexive controls only. Ultimate decidability depends also on evolutionary controls, aesthetic controls, and on an intuitively established and evolutionary validated set of norms. Time does not permit discussion of them here. Their introduction and application merely serve to support the statement that extrospection is only a part of a concept, indeed, that the criterion of fact, although a necessary and desirable part of the rational process, must be imbedded for its understanding in the context of the metasystem. The nature of the evolutionary control is to insure fruitfulness, adaptability and survival of a concept as a workable construct. Formalizations which have inherent ambiguities must necessarily sooner or later reach a condition where decidability cannot be established; and they must surely fail. This does not imply that once a method of rational inquiry is devised which accomplishes decidability it can be expected to retain this property indefinitely.

Novelty is a characteristic of emerging concepts. Novelty will inevitably occur in the method of inquiry itself. The appearance of higher orders of abstraction will make necessary a re-establishment of cognitive control evolving through a repetitive cycle of ambiguity, undecidability, and finally the discovery of new rational principles.

CONCLUSION. The conclusion I wish to draw from these remarks is that knowledge depends as much upon intuition as it does upon extrospection and logic; and that these aspects are interdependent. I have hoped to make you aware of the implication that the nature of the rational act is much more complicated than heretofore supposed and that the simplistic views of cognition must irrevocably be discarded.

HOW TO DESIGN WAR GAMES TO ANSWER RESEARCH QUESTIONS

William L. Pierce
Research Analysis Corporation

INTRODUCTION. In war gaming, to produce data for analysis, the game itself and the forms of data extraction must be designed to give outputs conveniently usable in answering the specific research questions that the game seeks to solve. This paper presents methods developed with this purpose in mind and employed in a rigidly-assessed, manually-played war game. With a brief historical sketch of the uses of war gaming from the 19th century to the present as background, present war games are classified into three groups, manual, computer-assisted, and computer-programmed, and defined. The design of manually-played war games is then considered separately in the context of a laboratory research tool. Finally, the application of the design requirements is illustrated by a description of TACSPIEL, RAC's division level war game.

Historically, war games have been developed and played to train officers and to test war plans. The former purpose was evident in the 9th century when Rigid Kriegspiel and Free Kriegspiel were developed. The testing of war plans by war games was used extensively by the Germans in the first half of the 20th century. After World War II, the technique of employing war games as an analytic tool was developed in an attempt to answer military questions pertaining to the battlefields of the future. With the advent of high speed computers, war gamers acquired a tool that permitted more comprehensive and complex games to be played. Also, the computer brought about a classification of war games by the war gaming community. A straightforward classification is to consider war games as either manually-operated, computer assisted, or computer programmed.

A manually-operated game is one in which all game orders are written, and assessments are made by people who are governed by strict or informal game rules, in essence, a rule book or an umpire.

A computer-assisted war game is a manually-operated game with the additional attribute that some of the bookkeeping and assessments are accomplished by a computer.

The third classification, the computer war game, is now in a prominent position in the war gaming field. In this type of game, after the start button is pushed, the computer plays the game without human intervention. Each and every situation thought to be important must be anticipated and simulated in the program with a suitable response.

THE DESIGN OF A WAR GAME AS A RESEARCH TOOL. Let us now consider a war game as a laboratory tool for military research. While much of what follows applies to all classifications of war games, I will be directing my words toward manual war games as a prelude to the later description of TACSPIEL.

One use of a laboratory tool is to enable the experimenter to investigate an area which would be inaccessible without the tool. For the military, the battle field of the future is the area at which attention is focused. To open this area for investigation, the war game becomes the tool. But for the experiments, or research plays, if you will, to be meaningful, the war game must be analytic. That is, it must be engineered to present a realistic environment for controlled experimental simulations with a view toward securing data for analysis.

It is not a difficult task for the military customer and the designer of war games to agree that war games can aid in the solution of military problems. However, when the research questions are directed at echelons from platoon to army, the designer must step back and take a sharp look at the design problems both obvious and subtle.

What does he see in the way of problems? First, there is the resolution problem. What is the gamet of resolution that should be considered in the game? Can the military units be played at company and battery level? Where and when can platoon, patrols, and radars be introduced? Do the research questions permit the game to be designed with divisions as the lowest echelon? To what resolution shall the unit deployments be recorded and played? How often should the game interactions be assessed?

The second problem which goes hand-in-glove with the resolution problem is the aggregation of the game models. If the basic unit is the company, then the models must reflect the capabilities of the company to move, fight, and receive casualties. There is a paramount requirement here when the designer builds the game models. Once he has chosen the resolution for the military units, he must be extremely careful to avoid constructing a game model for which no predictive data exists at the designed echelon. Should the input data to the model be lacking for the

echelon designed, the designer must re-examine the unit resolution. In short, resolution and aggregation as reflected in the game models are the two sides of the same coin.

Another problem which relates to the desired analytic nature of a manual war game is assessment of the play. A game can be played under a set of general rules with an umpire to assess battles, contacts, and other interactions. Or the game can be played under a set of rigid rules which are as detailed as the designer can make them. In this case, umpiring occurs infrequently and only when situations and capabilities arise that are not provided for in the rules. This latter method will produce the most objective and well-defined experimental conditions for a manually conducted game that can be attained.

Once the war gamer has decided on the resolution and aggregation level, he must now consider the basic tactical structure of the game. There are three characteristics which identify combat. They are the movement of units, the meeting of units known as contact, and the engagement by fire and maneuver of opposing units called battle. These characteristics are the basic tactical structure of a war game whether the game depicts ground, sea or air warfare. Any war game design must start by constructing models to represent these three characteristics.

Once the basic models of movement, contact, and battle exist, the war game is ready to consider the specific research questions of the military customer. When the research question is asked, the war gamer must ask himself three questions.

What models must be built such that the events to which the question is addressed will necessarily occur in the course of the play? What additional models must be designed to reflect the player's usual military capabilities, for example, artillery and tactical aircraft? How should all these models be constructed so that the output of each is presented in both usable form for analysis and with tactical realism for the players?

As an example relating to the first question, if the research question was to investigate the surveillance capability of a division in order to determine the detection rate of ground and airborne sensors, the war gamer would have to build, in detail, one model depicting the capability of each type of ground and airborne sensor including its associated delivery vehicle. These models would presumably allow the division commander as much flexibility as would be expected in actual combat and

would produce sufficient data for analysis. If the particular play was not directed at the surveillance question, an aggregate model depicting the intelligence acquisition capabilities of the airborne sensors could be built.

The third question concerning the presentation of the output of the models contains several requirements. The desired data for analysis must be presented in a format that can be easily manipulated manually or by a computer program. With the same format, the game assessments which contain the data for analysis should be presented unambiguously to the players and contain as much but no more information as could be expected to arise in actual combat under the same conditions that the model attempts to simulate. Without relaxing the above requirements, the recording of the assessments in a data format must not be time consuming. Otherwise the time saved during the analysis by preplanning the organization of the game data will be lost by the data recording process during the play of the game.

DESCRIPTION OF TACSPIEL. So far I have discussed the design of manually played war games pointing out the requirements for a basic tactical structure, associated sub-models, and data format.

I will now describe TACSPIEL, RAC's division-level war game, as an illustration of the application of the foregoing design principles and requirements.

The objective of TACSPIEL is ". . . to study operational problems of ground combat at division and lower echelon by analysis of play of a detailed tactical war game". The objective sets the framework within which the game was designed.

TACSPIEL is two-sided, free-play, analytic, rigidly-assessed, and manually operated. It is a free play game since after each side has been given their forces, scenarios, assigned missions, and approximate location of their respective reconnaissance elements, they are not constrained in their concepts of operation and organization other than the requirement to stay within the 45 x 200 km area of play.

It is analytic in that it is engineered to support research as a tactical environment for controlled experimental simulation of operational capabilities with a view toward securing data for analysis of their performance.

While the basic time resolution is one-half hour, battles are assessed on an hourly basis. That is, engagements are assessed once each hour of engagement. No winner or loser is declared but rather each force may accumulate casualties and the battle location can change if the attacker is successful. At the end of each hour of battle, the commanders receive reports of their own casualties, movement of the forces, and an estimate of casualties inflicted on the enemy. At that time they may attempt to reinforce or withdraw engaged units.

In the assessment of a battle, the engaged units basic combat effectiveness, their casualties at the start of the battle hour and the amount of casualties caused by enemy supporting artillery are used to calculate an attacker to defender force ratio. This force ratio is then used to determine by random number selection that hour's battle casualties on the attacker and defender, and the penetration of a successful attacker.

<u>ARTILLERY MODEL.</u> Since artillery has a capability which, in the real world, the division commander is able to employ with a high flexibility and effectiveness, the artillery model must be built to realistically reflect this capability. The emplacement time, rate of fire of the weapons, and the availability of the ammunition are the limitations on the employment of artillery. Since the effectiveness of rounds of different caliber to produce casualties vary, a standard unit of effectiveness called the fire unit or FU is used. One FU is equivalent in casualty production to 24 105-mm Howitzer HE rounds. The effectiveness of rounds of other calibers is equated to this measure. Thus, all fire missions are described and assessed, and a battery's basic load and resupply, computed in fire units.

The assessment of casualties is based on the number of fire units delivered, the type of target (armored, personnel) and posture of the target (exposed, attacking, defending, in woods, etc.) and the extent of observation on the target.

The results of a fire mission are reported to the side firing and the side receiving the shelling. The number of fire units expended, the type of target fired upon, and the target's location, are reported from the firing battery. If the fire is observed, an observer's report will contain the type and location of the target, an estimate of the damage inflicted, and the firing battery's designation to indicate what fire mission was being observed. For the unit receiving the fire, a shell report is generated containing an estimate of the amount of fire received and the amount of casualties suffered.

AIR OPERATIONS MODEL. The air operations model is designed to include the employment of air defense artillery, tactical aircraft against ground targets in support of engaged troops and against ground targets behind the enemy lines, airlift capability for divisional troops, and organic helicopters used in a reconnaissance role or to deliver fire from the air.

The effectiveness of the tactical aircraft's ordnance is equated to the artillery fire unit. Three stages of tactical air alert are played: No Alert, Standby, and On-station CAP. Three types of air missions are played, specific target, armed reconnaissance, and battle support. Appropriate planning times are assessed against aircraft when ordered from one of the air alert stages into one of the air missions.

Air defense artillery kill probabilities are based on their rate of fire, engagement ranges, altitude and speed of the aircraft.

The output of the air operations model is a report indicating the number of aircraft in the mission, the number surviving, and the result of the mission.

GROUND AND AIR SURVEILLANCE MODELS. To reflect the division's surveillance capability, OPs, patrols, and surveillance radars for the detection of moving personnel and vehicles are played in the ground surveillance model. The characteristics of the radars are obtained from the results of field tests.

The reconnaissance and surveillance capabilities of several airborne devices and agencies are amenable to war game simulation. For the purpose of TACSPIEL, however, only those systems designed to concentrate in the 10-km area immediately beyond the line of contact are played.

Based on the previous plays in which each surveillance mission beyond 10-km from the LC was individually played, an aggregate effectiveness has been developed for the surveillance capability in the zone in excess of 10 km beyond the LC. This aggregated deep penetration surveillance includes information gathered by air photos, infrared devices and side-looking air-borne radars.

The output of both the ground and air surveillance models reflects the normal capabilities of each sensor.

LOGISTICS AND VEHICLE BREAKDOWN MODELS. The next model illustrates how TACSPIEL was applied to generate data to support analysis of a research question. The research question concerned an analysis of consumption and resupply in the ROAD Division of Class III and V Supplies. To generate the data, a tactical logistics model was developed.

The model assumed an infinite stock of ammunition at the Army Supply Point. Using the organic transportation available in the division, the player was required to order up the ammunition he needed under a side condition that the fuel for the divisional units must be hauled simultaneously out of the same transportation capability.

All basic loads and ordering of ammunition were reduced to one unit, the "fire unit" of effect. The POL consumption rates and basic loads of the various units were also reduced to a single "consumption unit" or CU, equal to 18 gallons of gasoline. Finally, the transport available to haul the basic loads and resupply Class III and V was reduced to a "transportation serial" equivalent to 7 1/2 tons of lift. In this manner, players could requisition ammunition, POL, and transport in a system of units that was independent of the detailed tables of equipment of the organization concerned.

TACSPIEL has undertaken on a trial basis a model to simulate vehicle maintenance and mechanical failure of combat vehicles. Vehicle maintenance in the division is simulated by assuming that the level of availability of wheeled vehicles is in a steady state during the play of the game. The level of availability is assumed to be 80 percent for units having five or more trucks over one ton carrying capacity. This 20 percent loss of hauling capacity is reflected through reduction in basic loads of Class III and V supplies available to the player, and in resupply capabilities.

The breakdown model has been developed from data from field trials on the M60 tank and M113 APC. This model simulates mechanical failure of APC, armored vehicles and SP artillery. During each interval in which these types of units move, the unit is assessed for breakdown. If breakdown occurs, the unit effectiveness is degraded 5 percent. At appropriate times during the game, vehicles repaired at the divisional support group are returned to the game.

By adding new simulation models and proving them out TACSPIEL can expect to increase its potential to produce useful data for research.

MECHANIZATION OF TACSPIEL. The method by which game data are made available for analysis is by employing IBM punch cards as a medium of exchange for the vast bulk of orders and reports and for the recording of data. A vocabulary of codes has been developed to transmit the game messages between the players and Control. The orders recognized by the game rules and the assessments are punched on IBM cards using IBM port-a-punch holders and 40-column partially preperforated cards. These 40-column cards are divided into several fields. One format is used for player orders and another for assessments (Figure 2). These formats are designed to include a three-digit order code or report code, the unit's order of battle, its coordinates, and all pertinent information in the order or assessment.

The flow of orders from the players to Control, the assessment of the interactions pursuant to the orders, and the reporting of the results to the players is called the Order-Assess-Report Cycle (Figure 3).

The player-commanders write their mission orders and organize their units for combat using overlays and mission order forms, (Figure 4). The written orders are translated into TACSPIEL order codes in the ORDERS column of this form (Figure 5). After coding the orders, the orders are punched on the 40-column IBM cards, using the ORDER format. These cards then go to Control with their ground mission order.

Upon receipt by Control of the IBM cards, the data on these cards are transferred to standard 80-column IBM cards by an IBM Summary Punch. The 80-column cards are then used to prepare a worksheet for the assessors called the Unit History Form. The fòrmat (Figure 6) groups the combat units with their initial coded orders as they are organized for combat on the mission order to provide continuity in time and to help organize the assessorls work. Additional headings (ORDer, ACTion, POSition) are printed to permit the assessors to enter interval by interval any changes in orders from the players and assessment notes for each unit. Enough space on each page is available for listing units that become attached to an organization during the game. When the first page is filled, additional pages are produced which may reflect changes in the make-up of the combat organization.

After the Unit History Forms have been prepared, the action of the opposing forces is assessed. Reports generated by interactions are punched on the IBM cards in the Assessment Format, transferred by the IBM Summary Punch to standard IBM cards containing prepunched

military English, listed, and distributed to the players. For example, the output of an artillery assessment would use Code 801 for the firing unit's report and Code 750 for the observer's report, and the assessment by Control would be listed from the IBM cards (Figure 7). The percent casualties to the enemy unit and its order of battle would be deleted from the player's copy.

The player's response to the reports results in new orders by which a new Order-Assess-Report cycle is generated. Unless the general mission of a combat team is changed, the mission order form is not required for transmitting additional orders affecting that combat team.

By the use of the IBM cards and the mission order forms, a complete rapidly accessible record of the game is available for analysis. In addition, by the expedient of reproducing the cards, the data become accessible to any qualified study outside the gaming group itself.

SUMMARY. To summarize, in the design of a war game, a basic operational structure of movement, contact, and battle is required. Within this structure detailed simulations of the real world events to be studied are introduced in order to generate data to answer the specific research questions of the military customer.

In order to extract the data to answer the research questions rapidly and efficiently, the TACSPIEL war game has developed a method which combines a vocabulary of order and assessment codes with IBM cards. The result is a compact and complete game record for analysis and a data source on which analytic research can be based.

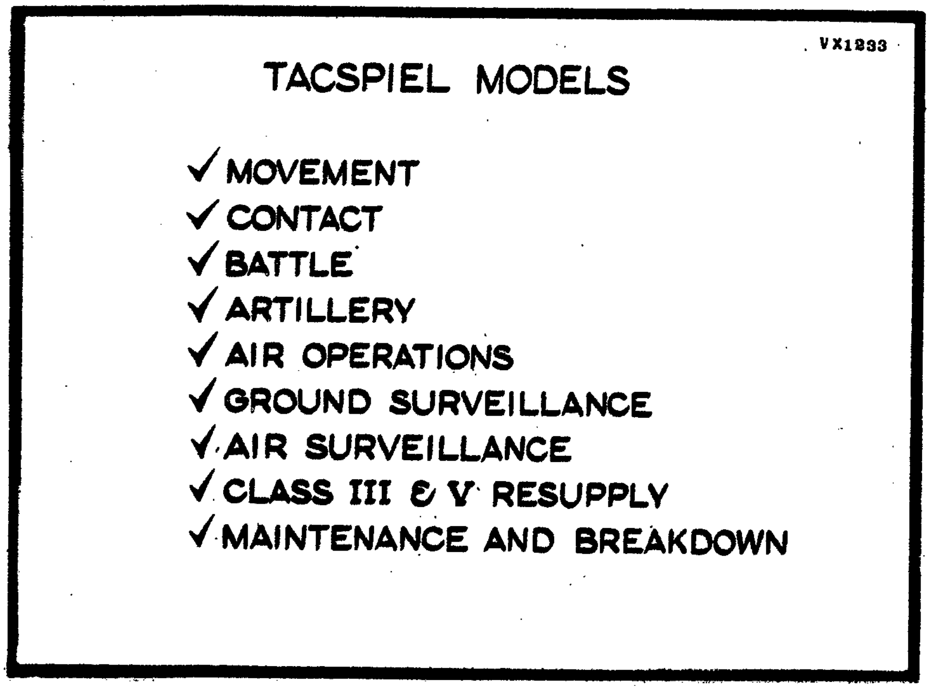


FIGURE 1- Tacspiel Models

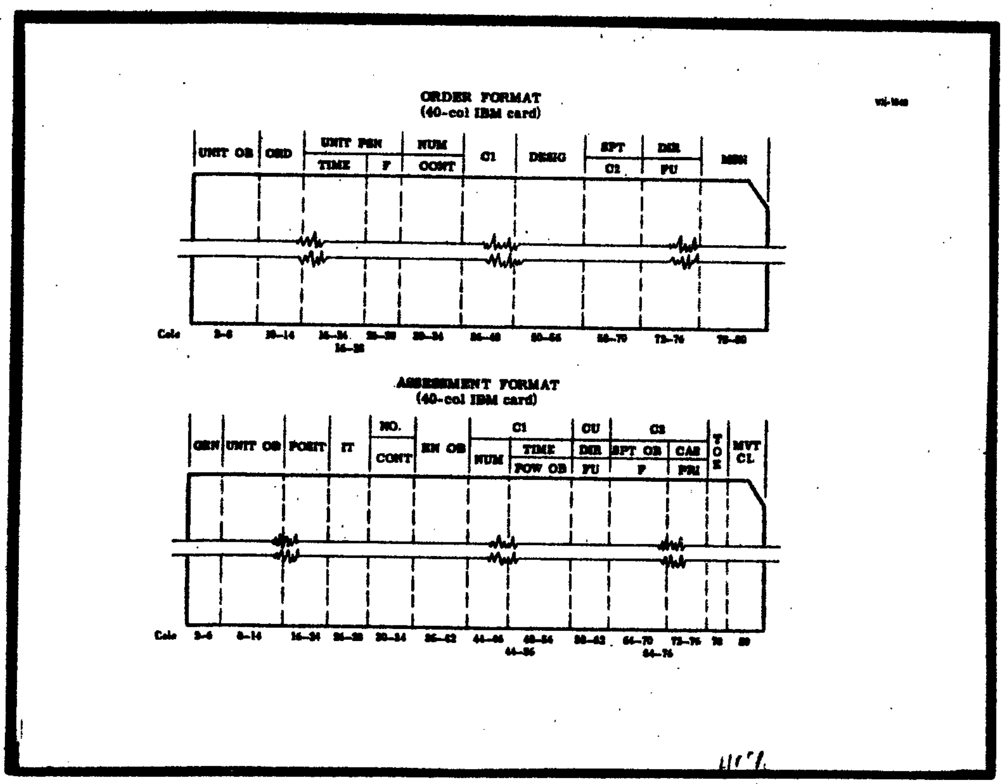


FIGURE 2- Order and Assessment Formats

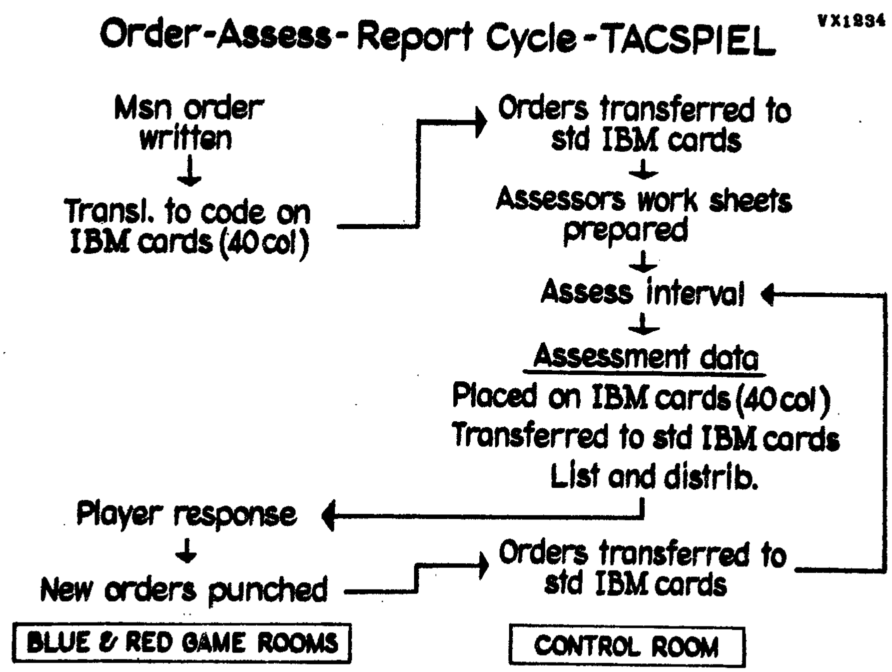
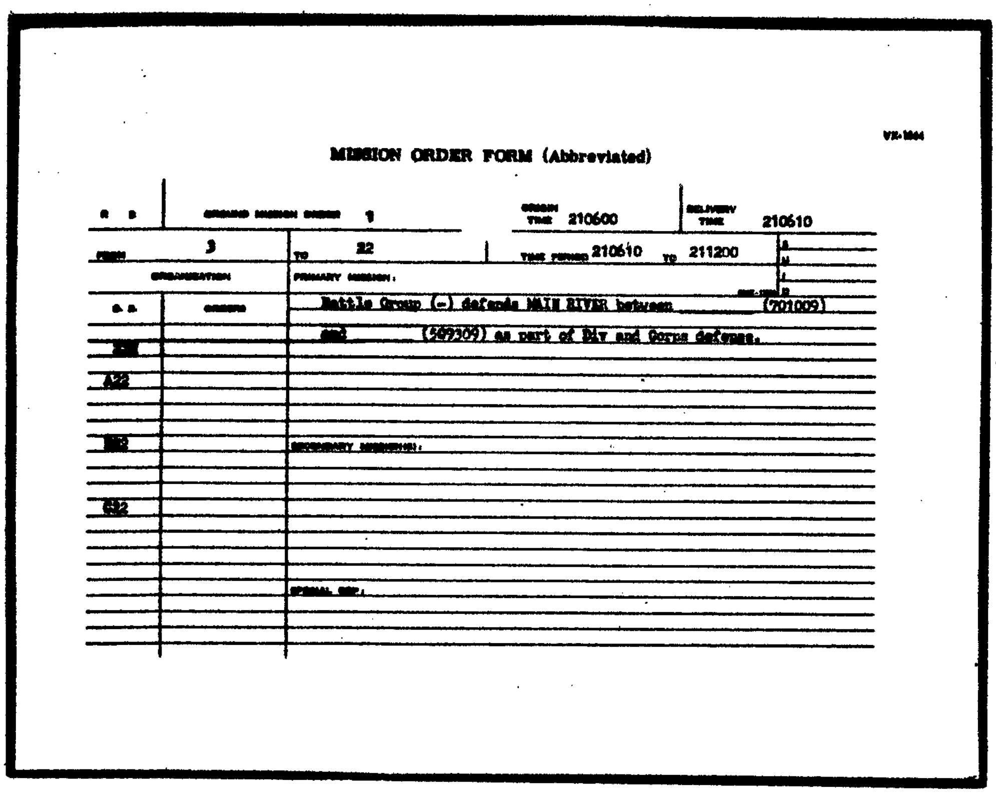


FIGURE 3- Order-Assess-Report Cycle


MISSION ORDER FORM (Abbreviated)

210600 210610

3 22 210610 211200

Battle Group (-) defends MAIN RIVER between (701009)

(549309) as part of Div and Corps defense.

FIGURE 4- Mission Order Form(Abbreviated)

This Page Intentionally Left Blank

MISSION ORDER FORM (Abbreviated)
with Coded Orders Added

VX-1846

R B | GROUND MISSION ORDER 1 | ORIGIN TIME 210600 | DELIVERY TIME 210610

FROM 3 | TO 22 | TIME PERIOD 210610 TO 211200

ORGANIZATION		
O. B.	ORDERS	PRIMARY MISSION:
		Battle Group (-) defends MAIN RIVER between (701009)
		and (509309) as part of Div and Corps defense.
22H	004 22	
A22	033 1	
	020 701009	
	507305	
B22	033 1	SECONDARY MISSION(S):
	020 700005	
	509309	
C22	032 1	
	020 690003	
	509307	
	131	
		SPECIAL SOP:

004 Assigned command of (designated organization) DESIG
020 Assigned zone/sector of responsibility C1 to C2
032 Assigned reserve role in MSN
033 Assigned forward defense role in MSN
131 Defend assigned sector

FIGURE 5- Mission Order Form (Abbreviated) with Coded Orders Added

VX-1846

R B GAME ______ UNIT HISTORY FORM PAGE ___ OF ___ TIME ___ TO ___

CMD ECH	UNIT	ORD	UNIT TIME	PSN F	NUM CONT	C1	DESIG	SPT C2	DIR FU	MSN	TIME: ACT POS	TO ORD ACT POS
22BG												
	22H	004					22					
	A22	033								1		
		020				701009		507305				
	B22	033								1		
		020				700005		509309				
	C22	032								1		
		020				690003		509307				
		131										

FIGURE 6- Unit History Form

This Page Intentionally
Left Blank

ARTILLERY ASSESSMENT CODES VX-1647

801 SPECIFIC FIRE MISSION:

- Expended FU Fire Units
- On CONT-type tgt at C1
- F-type Ammo
- % caus to tgt-CAS
- En Order of Battle-ENOB

750 OBSERVER REPORT:

- Est. damage fr F (use OB of arty unit)
- to en tgt CONT at C1
- in FU (n, l, m, h)
- En Order of Battle-ENOB

R GEN	B UNIT	POSIT	IT	ASSESSMENT REPORT							TIME ____ TO ____
801	A12	7099	812	TGT AT	6098	RECD	10	FU,	.05	CAS	A61
750	B77	6197	812	VICINITY	6098	DAM REC	L	FR	A12		A61

Note: Underlined items represent data punched from assessment.

FIGURE 7- Artillery Assessment Codes and Report

This Page Intentionally
Left Blank

EVALUATION OF PERFORMANCE RELIABILITY USING REGRESSION MODELS

Seymour K. Einbinder, Picatinny Arsenal
and
Ingram Olkin, Stanford University

0. ABSTRACT. Performance reliability is the probability that a weapon will perform its prescribed function under given conditions of environment at some particular time. Performance reliability models are defined for continuous performance response variables. Procedures are then described for the evaluation of reliability with emphasis on the application of univariate and multivariate regression analysis to single and multiple continuous response variables, respectively. Point and confidence interval estimation methods for performance reliability are discussed, and a sample problem is presented illustrating some of the basic concepts and results.

1. INTRODUCTION. A major problem during the research and development of a weapon or warhead is the assurance of high functioning reliability of the final prototype design. The reliability concepts and evaluation methods to be described are general and are applicable to a wide variety of systems and components.

A weapon during its lifetime may be subjected to many environmental factors or stresses such as temperature, vibration, acceleration, rough handling, etc. In addition, the stresses may be encountered singly, simultaneously or in sequence. The problem of testing and estimating reliability is of importance to the weapon developer in order to assure the user of a reliable weapon for use in any potential combat situation.

The establishment of high reliability with a high level of confidence generally requires the testing of larger numbers of items than are usually available during a development program for a complex and expensive item. Thus, it is generally necessary to obtain the most information with a minimum number of samples and tests. Improved and more efficient statistical methods are required in many cases to solve the reliability estimation problem.

Before solutions to a problem can be obtained, it is important to delineate the problem so that a representative mathematical model can be developed. Obviously, any solutions obtained can be no better than the underlying mathematical model which is assumed as a prototype of the problem.

In a previous paper [3], the emphasis was on test-to-failure and stress vs. strength analyses for single and multiple environments (stresses). The present paper is concerned mainly with reliability in the case of a continuous regression response surface, and thereby is an extension of [3]. Methods of point and interval estimation for the univariate and multivariate regression problems are discussed.

2. RELIABILITY CONCEPTS. Reliability of a weapon may be defined as the probability of a successful functioning under required conditions of the environment at some particular time. Successful functioning might require the successful operation of several or all components of a system. The outputs or responses of each component may be attribute or continuous. In the case of the continuous response, success may require that the response lie within certain limits (possibly specification limits),

To illustrate these concepts, a hypothetical shaped charge warhead section for a missile will be used as an example. Successful functioning of the warhead section may require that the S and A (Safety and Arming Device) must arm and detonate the warhead on target impact, and the warhead must then penetrate at least a specified distance into an armor plate target. This example could be further complicated by specifying arming limits for the S and A. Failure of the warhead section may occur in two fundamental ways:

(1) a complete dud or catastrophic failure may result such that no warhead detonation takes place, or,

(2) the warhead explosive train may be initiated but the armor penetration requirement may not be met.

The reliability of the warhead section is given by

$$R = (1 - P_D)\, R_{WHD},$$

where R is the overall warhead section reliability, P_D is the probability of a dud or catastrophic failure, and R_{WHD} is the conditional probability that the warhead exceeds the specified performance requirement. The latter will be referred to as the performance reliability and is of prime concern in this paper. The dud probability can be broken down

further according to various components. In general it is necessary to evaluate the dud or catastrophic failures separately from the performance failures since they do not have the same distribution and are mutually exclusive. Dud failures, being attribute, normally require larger sample sizes for evaluation with the same precision and confidence levels as would performance reliability based on continuous variables.

The remainder of this paper will be concerned with the evaluation of the performance reliability, R_{WHD}.

3. MATHEMATICAL MODELS. In this section we define the mathematical models upon which subsequent analysis is based. Univariate and multivariate responses and single and multiple stresses are considered. Estimation procedures are described in the subsequent sections.

3.1 Univariate Response. We have defined performance reliability as the probability that a continuous performance variable lies within certain specified limits. Thus, it may be required that the arming time for an S and A Device be greater than some minimum time required for safety. The performance variable or response may be thought of as a dependent variable which is a function of one or more environments or stresses which are the independent variables. In general, this functional relationship is unknown; however, we can approximate the response function over small regions of the function space by linear regression methods. Generally, we are concerned with the reliability under some critical stress conditions. These conditions will be referred to as a critical point or critical reliability boundary. The regression experiment is designed to provide the best information in the vicinity of this point. Figure 1 illustrates the experimental design in general terms for the univariate case. $(x_1, \ldots, x_m)$ represent the applied stresses or environments such as temperature, vibration, etc., and the elements of the design matrix represent the levels of each of these stresses. For example, x_{12} represents the second level of the stress x_1, etc. The column titled response vector represents the observed response obtained with each treatment combination. The response, y, is a continuous variable such as arming time in the case of the fuze, or possibly depth of penetration in the case of a shaped charge warhead; the response, y_i for the i^{th} treatment combination may be expressed as a linear combination of the treatment levels plus some random error. The regression model and underlying assumptions are:

$$y = X'\beta + u,$$

y:	$n \times 1$	is the observation vector,
u:	$n \times 1$	is the vector of random errors, and is normally distributed with mean vector O and covariance matrix $\sigma^2 I$, i.e., $E(u) = O$, $E(u u') = \sigma^2 I$,
X:	$m \times n$	is the design matrix of rank r, $r \leq m \leq n$,
β:	$m \times 1$	is the vector of regression coefficients.

A geometrical interpretation of this regression model and its relation to performance reliability is shown in Figure 2 [Tables and figures can be found at the end of the article.] for the univariate response case. $(x_1, \ldots, x_m)$ are the stress variables, $(c_1, \ldots, c_m)$ are the components of the critical reliability boundary vector c. The points shown in the $(x_1, \ldots, x_m)$ plane represent the treatment combinations for the regression experiment, and the average response, y, is represented by the regression surface shown above the points. The regression equation provides estimates of the response for any point in the $(x_1, \ldots, x_m)$ space. The response y at some point such as the critical reliability boundary $c: m \times 1$ is denoted by $y^{(c)}$ and is distributed according to $N(c'\beta, \sigma^2)$. The lower performance limit which the response $y^{(c)}$ is required to exceed is denoted by $y^{(O)}$, and consequently, the performance reliability R is given by:

$$(1)\quad R(c) = P\{Y^{(c)} \geq y^{(O)}\} = \int_{y^{(O)}}^{\infty} n(y^{(c)} \mid c'\beta, \sigma^2)\, dy^{(c)} \equiv g_1(c'\beta, \sigma^2; c),$$

where $n(x \mid \mu, \sigma^2)$ is the density function for a normal population with mean μ and variance σ^2. This expression represents the shaded area under the normal curve shown in Figure 2.

Thus, our problem is to estimate g which is a function of the unknown parameters β, σ^2, and the fixed point c, based upon a sample of size n treated as a single or multiple regression experiment.

3.2 Multivariate Response. The univariate model will now be extended to include cases where more than one continuous response may be observed on a single experimental unit such as S and A arming time, functioning time, and self-destruct time; also, the responses may be correlated. Multivariate analysis techniques permit the correlation between responses to be investigated. As before, the problem is best illustrated by examining the table shown in Figure 3. The design matrix X is exactly the same as for the univariate case; $(x_1, \ldots, x_m)$ is still the vector of applied stresses. However, instead of a single response vector of y's, we now have p responses $(y_1, \ldots, y_p)$. Thus, for each treatment combination we observe p responses so that our response vector for the univariate case has now become a response matrix where the column vectors may be correlated, and the rows which represent independent response vectors are uncorrelated. The multivariate model and assumptions are:

$$Y = X'B + U,$$

Y: $n \times p$	is the response matrix
X: $m \times n$	is the design matrix of rank $r \leq m < n$, $p \leq n-r$,
B: $m \times p$	is the matrix of regression coefficients,
U: $n \times p$	is the error matrix,
u_j, $j=1, \ldots, n$	are the rows of U and are independently and identically distributed, each having a p-variate normal distribution with mean vector O and positive definite covariance matrix Σ.

From the multivariate model, a $p \times 1$ response vector $y^{(c)}$ is obtained for the response at the critical reliability boundary vector c: $m \times 1$. The response vector $y^{(c)}$ is distributed according to $N(c'B, \Sigma)$, in which the $p \times p$ covariance matrix Σ takes into account any correlations between responses.

The performance reliability R for the multiple response case is given by:

$$(2)\quad R(c) = P\left\{Y_1^{(c)} \geq y_1^{(O)}, \ldots, Y_p^{(c)} \geq y_p^{(O)}\right\}$$

$$= \int_{y_1^{(O)}}^{\infty} \cdots \int_{y_p^{(O)}}^{\infty} n(y^{(c)} \mid c' B, \Sigma)\, dy_1^{(c)} \ldots dy_p^{(c)}$$

$$= g_p\,(c' B,\ \Sigma;\ c)\ .$$

A graphical representation of the performance reliability in two dimensions is shown in Figure 4 for the multivariate case. The above integral represents the volume of the multivariate normal density function over the shaded quadrant whose vertex $y^{(O)}$ is the vector of specification limits.

Thus, the general problem may be summarized as follows: Based upon the results of a suitable experimental design with a sample of size n, it is required to estimate the g function for the univariate and multivariate cases, both by point estimation and confidence limits.

4. POINT ESTIMATION. The general problem is to estimate the performance reliability functions defined for the univariate and multivariate responses both by a point estimate and confidence limits based upon responses observed on a sample of size n subjected to various stress treatments in accordance with a suitable experimental design. The experimental designs used for exploring response surfaces [1, 2] are generally suitable for exploring the region around the critical reliability boundary.

4.1 Univariate Point Estimates. The g or R functions to be estimated may be written as follows for the univariate case:

$$R(\beta, \sigma^2) \equiv g_1(c'\beta, \sigma^2;\ c) = \int_{(y^{(O)}-c'\beta)/\sigma}^{\infty} (2\pi)^{-1/2} \exp\,(-t^2/2)\, dt,$$

where c and β are $m \times 1$ vectors.

We consider three types of point estimates. Suppose we write

$$K(\beta, \sigma) = (y^{(O)} - c'\beta)/\sigma \quad ,$$

then

$$(3) \qquad R(\beta, \sigma^2) = \int_{K(\beta,\sigma)}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\, dt.$$

The first estimate of R is based on using $K(\hat{\beta}, \hat{\sigma})$, where $\hat{\beta}$ and $\hat{\sigma}$ are appropriate estimates of β and σ.

The least squares estimate of β is given by

$$(4) \qquad \hat{\beta} = (XX')^{-1}X\, y \, ,$$

where X: $m \times n$, of rank $m \leq n$, is the design matrix, and y : $n \times 1$ is the response vector. An unbiased estimator of σ^2 is

$$(5) \qquad \hat{\sigma}^2 = \frac{(y-X'\hat{\beta})'(y-X'\hat{\beta})}{n-m} \quad .$$

Thus, we may use the estimate

$$(6) \qquad K(\hat{\beta}, \hat{\sigma}) = (y^{(O)} - c'\beta)/\hat{\sigma} \quad ,$$

from which we obtain

$$(7) \qquad R(\hat{\beta}, \hat{\sigma}^2) = \int_{K(\hat{\beta}, \hat{\sigma})}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\, dt.$$

A second estimate is based on the UMVU estimate of K, namely

(8) $$\tilde{K}(\beta,\sigma) = \sqrt{\frac{2}{f}}\,\frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\,\frac{y^{(0)} - c'\hat{\beta}}{\hat{\sigma}} = \sqrt{\frac{2}{f}}\,\frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\,K(\hat{\beta},\hat{\sigma}) \quad ,$$

where $f = n-m$, from which we may use

(9) $$\tilde{R}(\beta,\sigma^2) = \int_{\tilde{K}(\beta,\sigma)}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\ dt$$

as an estimate at $R(\beta,\sigma^2)$.

Although $\tilde{K}$ is a UMVU estimate of K, it is not the case that $\tilde{R}$ is a UMVU estimate of R. Consequently, a third procedure is based on the UMVU estimate of R, [4], and is given by

(10) $$\bar{R}(\beta,\sigma^2) = \int_0^{\max[0,\eta]} \frac{t^{\left[\frac{f-1}{2} - 1\right]}(1-t)^{\left[\frac{f-1}{2} - 1\right]}}{B\left(\frac{f-1}{2},\frac{f-1}{2}\right)}\ dt \ ,$$

where $\eta = \frac{1}{2} - \frac{(y^{(10)} - c'\beta)}{2\hat{\sigma}\sqrt{f(1-c'(XX')^{-1}c)}} = \frac{1}{2} - \frac{K(\hat{\beta},\hat{\sigma})}{2\sqrt{f(1-c'(XX')^{-1}c)}}$.

Note that $\bar{R}(\beta,\sigma^2) = 1$ if $\eta > 1$, and that the estimate is valid for critical vectors c such that $c'(XX')^{-1}c < 1$.

Unfortunately, comparisons of the risk of these estimators are unavailable, since the determination of the variance is quite complicated, and was not attempted in this paper.

4.2 Multivariate Case. In the multivariate case, we have

$$R(B;\Sigma) = \int_{y_1^{(O)}}^{\infty} \int_{y_p^{(O)}}^{\infty} \frac{e^{-1/2 \ \mathrm{tr}\ \Sigma^{-1} (Y-X'B)'(Y-X'B)}}{|\Sigma|^{pn/2}(2\pi)^{pn/2}} \ dY,$$

where $Y : n \times p$, $X : m \times n$, $B : m \times p$. As in the univariate case, we can consider $R(\hat{B}, \hat{\Sigma})$ as an estimate of $R(B, \Sigma)$, where $\hat{B} = (XX')^{-1}XY$, and $\hat{\Sigma} = (Y - X'\hat{B})'(Y - X'\hat{B}) / [p(n-m)]$. The problem, however, is still to evaluate the multivariate normal distribution over an orthant. In fact, whether we use this estimation procedure or another, the difficulty of carrying out such an integration still remains. However, for any particular problem, one can employ numerical procedures to yield an answer. Another possibility which has not been considered in the literature is to obtain a lower bound for $R(B, \Sigma)$ in terms of known functions. Further work in this area is required.

5. CONFIDENCE INTERVALS. The problem of obtaining confidence intervals for the g or R functions is considered next. The general method is discussed in [3], and is now extended to the regression model. In Section 4, three estimates were presented. For only the second procedure is the distribution theory known, so that exact confidence intervals can be obtained. However, the first procedure does lead to approximate or asymptotic intervals based on the normal distribution.

5.1 Exact Confidence Intervals. Since $R(\beta, \sigma^2)$ is a monotone function of $K(\beta, \sigma)$, if we can find a confidence interval (K_1, K_2) for K, we will then have a confidence interval (R_1, R_2) for R, where

$$R_i = \int_{K_i}^{\infty} (2\pi)^{-1/2} \exp(-y^2/2)\, dy.$$

It is shown in Appendix A that $K(\hat{\beta}, \hat{\sigma})/\|a\| \equiv t(f, \delta)$, where $\|a\|^2 = c'(XX')^{-1}c$, has a non-central t-distribution with $f = n - m$ degrees of freedom and non-centrality parameter.

$$\delta = (y^{(O)} - c'\beta)/(\sigma \|a\|) = K(\beta, \sigma)/\|a\| \quad .$$

Thus, a lower and upper confidence limit with confidence coefficient $1 - \alpha$ may be obtained by finding the values of δ_i for which

(11) $P\left\{t > \frac{K(\hat{\beta}, \hat{\sigma})}{\|a\|} \,\Big|\, f, \ \delta_i\right\} = \alpha_i, \quad i = 1, 2,$ where $\alpha_1 = 1 - \alpha/2$, and

$\alpha_2 = \alpha/2$. Table IV in [6] may be used to obtain δ_i for seventeen values of ϵ.

The tabulation by Resnikoff and Lieberman [6] of the percentage points of the non-central t-statistic may be conveniently used to obtain the limits δ_1 and δ_2 that satisfy (11). The entries in the table give the values of x such that

$$P\left\{\frac{t}{\sqrt{f}} > x\right\} = \epsilon \quad .$$

The table should be entered for the degrees of freedom $f = n-m$, the probability ϵ_i corresponding to α_i, and $x = K(\hat{\beta}, \hat{\sigma})/(\|a\|\sqrt{f})$. The required non-centrality value $\delta_i = \sqrt{f+1}\ K_p$, where K_p is the standardized normal random variable exceeded with probability p. The present concern was with one sided tails (one sided specification limits) for both the univariate and multivariate cases. A review of available point and confidence methods for two sided tails is given in [3].

5.2 Approximate Confidence Intervals. If we expand $R(\hat{\beta}, \hat{\sigma}^2)$ about $R(\beta, \sigma^2)$, we obtain the result that

$$R(\hat{\beta}, \hat{\sigma}^2) - R(\beta, \sigma^2) \sim N(O, V_\infty(\beta, \sigma^2)) \quad ,$$

where

$$V_\infty(\beta, \sigma) \equiv \sigma^2 \left[n(y^{(O)} | c'\beta, \sigma^2)\right]^2 \left\{c'(XX')^{-1} c + \frac{(y^{(O)} - c'\beta)^2}{2\sigma^2 f}\right\} .$$

Consequently (see Appendix B),

$$\frac{R(\hat{\beta}, \hat{\sigma}^2) - R(\beta, \sigma^2)}{\sqrt{V_\infty(\beta, \hat{\sigma})}} \sim N(O; 1) \quad ,$$

from which we obtain the confidence interval

$$\left[R(\beta, \sigma^2) + z_{\alpha/2}\sqrt{V_\infty(\hat{\beta}, \hat{\sigma})}\ ,\ R(\hat{\beta}, \hat{\sigma}^2) + z_{1-\alpha/2}\sqrt{V_\infty(\beta, \sigma)}\ \right],$$

where z_α is the 100 α % point of the N(O; 1) distribution.

6. SAMPLE PROBLEM. In order to illustrate the results of the previous sections, a sample problem will be solved. A model representing the performance of a hypothetical shaped charge warhead section for a missile will be described, and the performance reliability will be evaluated based upon a Monte Carlo simulation of test results. The point estimates and confidence intervals obtained using the methods previously described will be compared with the true value of the reliability. Only the univariate or independent response cases are considered.

6.1 Performance Model. The warhead section to be evaluated consists of a shaped charge warhead and a Safety and Arming Device. It is assumed that the warhead is required to penetrate at least 10 inches into an armor plate target and that the minimum arming time of the S and A is O. 5 seconds. The warhead section is expected to meet these performance requirements under all possible combinations of vibration and temperature shock that may be encountered. To facilitate the illustration, only two stresses are considered in this problem, but the procedure is easily extended to more than two stress variables.

The two stresses, vibration in g's and temperature shock in standard cycles, are denoted by X_1 and X_2, respectively. Coded levels of the stresses are used throughout this problem to facilitate the analysis and simulation of test results. The relationship between the coded and actual stress units is of no importance with regard to illustrating the reliability evaluation methods and will be disregarded.

The critical reliability boundary is defined by the vector $c' = (c_0, c_1, c_2)$ where c_1, c_2 are the upper stress limits specified for vibration and temperature shock, respectively and c_0 is a dummy variable required to make the vector c consistent with the design matrix X and is equal to 1. The coded variables, in this example, are centered on the critical reliability boundary so that $c' = (1, 0, 0)$.

The warhead performance is measured in terms of depth of penetration t_w into monolithic armor, and S and A performance is measured by arming time t_f. The distribution of warhead penetration t_w and arming time t_f for the S and A is each distributed according to $N[\beta_0 + \beta_1 x_1 + \beta_2 x_2, \sigma^2]$. The true values of the parameters are

	β_0	β_1	β_2	σ
Warhead	13''	-0.6	-0.4	1.5''
S and A	0.6 sec.	0.07	0.03	.033 sec.

These models thus assume that the average penetration decreases linearly with increasing stress and that the average arming time increases very slowly with increased stress within the region of interest. Thus, by (1), we see that the performance reliability for the warhead and S and A, respectively, are

$$(12) \qquad R_{WHD} = P\{t_w^{(c)} \geq 10''\} = \int_{10''}^{\infty} n(t_w^{(c)} \mid c'\beta, \sigma^2)\, dt_w^{(c)}$$

$$= g_1(c'\beta, \sigma; c) \equiv g_1(13'', 1.5''; c) = 0.977$$

$$(13) \qquad R_{S \text{ and } A} = P\left\{t_f^{(c)} \geq 0.5\right\} = \int_{0.5}^{\infty} n(t_f^{(c)} \mid c'\beta, \sigma^2)\, dt_f^{(c)}$$

$$\doteq g_1(c'\beta,\ \sigma;\ c) \equiv g_1(0.6,\ .033\ ;\ c) = 0.999\ .$$

The performance reliability of the warhead section is thus

$$R = R_{WHD} \cdot R_{S \text{ and } A} = .976\ .$$

Dud probabilities were not considered in this model. The evaluation of dud rates requires attribute test methods which are not as efficient as the variables plans and require much larger sample sizes. In conducting the type of test program described herein an estimate of the dud reliability may be made by noting the number of dud failures. However, useful interval estimation with these results may not be possible with reasonable confidence coefficients. When a dud occurs, it is desirable to repeat the appropriate test under the same conditions in order to avoid or minimize having to work with missing data in the test plan.

6.2. Multiple Regression Analysis. Multiple linear regression experimental designs of the type used in exploring response surfaces were used to evaluate the performance reliability of the warhead and S and A based upon the stated performance model. In particular, central composite rotatable experimental designs [3] were used. The experiments were conducted with sample sizes (n) of 8 and 30 for both the warhead and S and A. The treatment combinations and the responses generated by Monte Carlo simulation of the performance models are shown in Tables 1 to 4.

Least Squares estimates of the regression coefficients and error variance were made for the test results, and goodness-of-fit tests were conducted. In all four cases, a linear regression model was found to represent the data adequately. The least squares estimates of the regression coefficients obtained for each case are as follows:

Item	n	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Warhead	8	13.99	-.175	.675
	30	12.75	-.975	.215
S and A	8	.59	.055	.020
	30	.60	,0705	.0205

Tests of significance at the .05 level performed for the regression coefficients gave the following results. The estimates of β_1 and β_2 for the warhead based on $n = 8$ were not significantly different from zero. For $n = 30$, $\hat{\beta}_2$ was not significantly different from zero, but $\hat{\beta}_1$ which corresponds to the effect of the vibration stress X_1 was found to be significantly different from zero, which corresponds to the true situation for the model. In the case of the S and A, $\hat{\beta}_2$(temperature shock) was not significantly different from zero, and $\hat{\beta}_1$ (vibration) was significantly different from zero for n=8 and 30.

Point estimates of the performance reliability $R(\beta, \sigma^2)$ at $c'=(1, 0, 0)$ were made using the UMVU estimate $\tilde{K}(\beta, \sigma)$ of $K(\beta, \sigma)$. Exact one sided lower confidence limits using the non-central t-distribution were also obtained. A summary of these results is tabulated below, and a sample computation is given in Appendix C for one case. Estimates of $R(\beta, \sigma^2)$ based on the estimates $K(\hat{\beta}, \hat{\sigma})$ and $\tilde{K}(\beta, \sigma)$ are also included in the appendix.

Case	Item	n	$\tilde{R}$	R(.95)	True R
1	Warhead	8	.974	.832	.977
2		30	.967	.905	
3	S and A	8	.983	.891	.999
4		30	.997	.981	

where $\tilde{R}$ is the Estimate of Performance Reliability based on the UMVU estimate $\tilde{K}(\beta, \sigma)$ or $K(b, \sigma)$ and R(.95) is the one-sided lower 95% conficence limit for Performance Reliability.

A point estimate of the warhead section reliability is given by

$$\tilde{R} = \tilde{R}_{WHD} \cdot \tilde{R}_{S \text{ and } A} .$$

Conservative .90 confidence intervals for the warhead section reliability are obtained by multiplying the lower .95 confidence limits for the warhead and S and A. This result is easily proven by applying the Bonferroni inequality to obtain a conservative simultaneous confidence region T for R_{WHD} and $R_{S \text{ and } A}$ and by making use of the fact that the product is monotone in each variable. Thus, we obtain the following results for the warhead section reliability.

n	$\tilde{R}$	R(≥ .90)	True R
8	.957	.741	.976
30	.964	.888	

REFERENCES

1. Box, G. E. P. and Draper, N. R., "A Basis for the Selection of a Response Surface Design", J. Amer. Statist. Assoc., 54 (1959) 622-654.

2. Box, G. E. P. and Hunter, J. S., "Multifactor Experimental Designs for Exploring Response Surfaces", Ann. Math. Statist., 28 (1957) 195-241.

3. Einbinder, S. K. and Olkin, I., "Reliability Testing and Estimation for Single and Multiple Environments (Preliminary Report)", Proceedings of the Seventh Conference on the Design of Experiments in Army Research Development and Testing, U. S. Army Research Office, Report No. 62-2, August, 1962, 261-291.

4. Ghurye, S. G. and Olkin, I., "Unbiased Estimation of Probability Densities", Abstract, Annals. Math. Statist., 32 (1961) 1346.

5. Johnson, N. L. and Welch, B. L., "Applications of the Non-central t-Distribution", Biometrika, 31 (1940) 362-389.

6. Resnikoff, George J. and Lieberman, Gerald J., Tables of the Non-central t-Distribution, Stanford University Press, Stanford, California, 1957.

TABLE I

WARHEAD

EXPERIMENTAL DESIGN N=8

Sample	x_1	x_2	t_w
1	-1	-1	13.3
2	+1	-1	14.9
3	-1	+1	16.6
4	+1	+1	14.3
5	0	0	12.6
6	0	0	12.7
7	0	0	12.2
8	0	0	15.3

x_1 = Vibration

x_2 = Temperature Shock

t_w = Penetration (inches)

TABLE 2

WARHEAD EXPERIMENTAL DESIGN N=30

Sample	x_1	x_2	t_w
1	-1	-1	13.8
2	+1	-1	12.1
3	-1	+1	13.8
4	+1	+1	12.7
5	0	0	12.1
6	0	0	10.9
7	0	0	12.5
8	0	0	12.2
9	-1	-1	13.7
10	+1	-1	10.5
11	-1	+1	12.7
12	+1	+1	11.0
13	0	0	14.8
14	-1	-1	13.7
15	+1	-1	9.5
16	-1	+1	11.5
17	+1	+1	13.0
18	0	0	13.3
19	-1	-1	12.0
20	+1	-1	12.5
21	-1	+1	14.4
22	+1	+1	11.4
23	0	0	13.9
24	-1	-1	14.8
25	+1	-1	11.5
26	-1	+1	15.6
27	+1	+1	14.3
28	0	0	12.0
29	0	0	16.0
30	0	0	12.3

x_1 = Vibration

x_2 = Temperature Shock

t_w = Penetration (inches)

TABLE 3

S & A EXPERIMENTAL DESIGN N=8

Sample	x_1	x_2	Arming Time (seconds)
1	-1	-1	.48
2	+1	-1	.63
3	-1	+1	.56
4	+1	+1	.63
5	0	0	.61
6	0	0	.56
7	0	0	.60
8	0	0	.64

x_1 = Vibration

x_2 = Temperature Shock

TABLE 4

S & A EXPERIMENTAL DESIGN N=30

Sample	x_1	x_2	Arming Time (seconds)
1	-1	-1	.55
2	+1	-1	.63
3	-1	+1	.56
4	+1	+1	.71
5	0	0	.57
6	0	0	.68
7	0	0	.67
8	0	0	.60
9	-1	-1	.50
10	+1	-1	.66
11	-1	+1	.53
12	+1	+1	.73
13	0	0	.57
14	-1	-1	.58
15	+1	-1	.66
16	-1	+1	.57
17	+1	+1	.72
18	0	0	.55
19	-1	-1	.46
20	+1	-1	.64
21	-1	+1	.54
22	+1	+1	.65
23	0	0	.59
24	-1	-1	.51
25	+1	-1	.67
26	-1	+1	.56
27	+1	+1	.70
28	0	0	.53
29	0	0	.61
30	0	0	.60

x_1 = Vibration

x_2 = Temperature Shock

UNIVARIATE RESPONSE

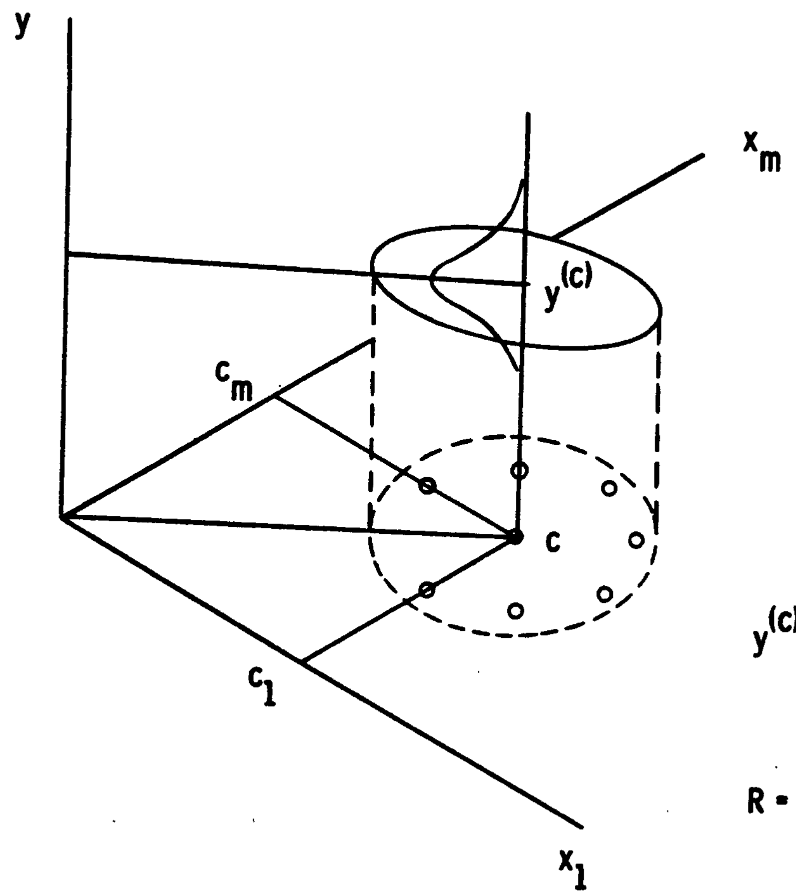
No.	Response Vector Y	Design Matrix					
		x_1	x_2	.	.	.	x_m
1	y_1	x_{11}	x_{21}	.	.	.	x_{m1}
2	y_2	x_{12}	x_{22}	.	.	.	x_{m2}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
l	y_l	x_{1l}	x_{2l}	.	.	.	x_{ml}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	y_n	x_{1n}	x_{2n}	.	.	.	x_{mn}

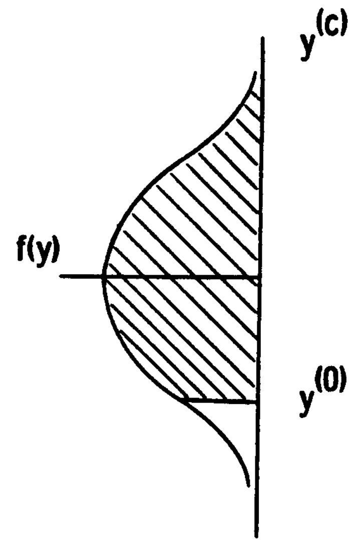
$$\Omega: \; y_l = \beta_1 x_{1l} + \beta_2 x_{2l} + \cdots + \beta_m x_{ml} + u_l, \; l = 1,\ldots,n$$

$\{u_1,\ldots,u_n\}$ are independently and identically distributed with mean 0 and variance σ^2.

Figure I

PERFORMANCE RELIABILITY
UNIVARIATE RESPONSE





$$y^{(c)} \sim N[c_1\beta_1 + c_2\beta_2 + \cdots + c_m\beta_m, \sigma^2]$$
$$= N[c'\beta, \sigma^2]$$
$$R = P\{Y^{(c)} \geq y^{(0)}\} = \int_{y^{(0)}}^{\infty} n(c'\beta, \sigma^2)\, dy^{(c)}$$

Figure 2

MULTIVARIATE RESPONSE

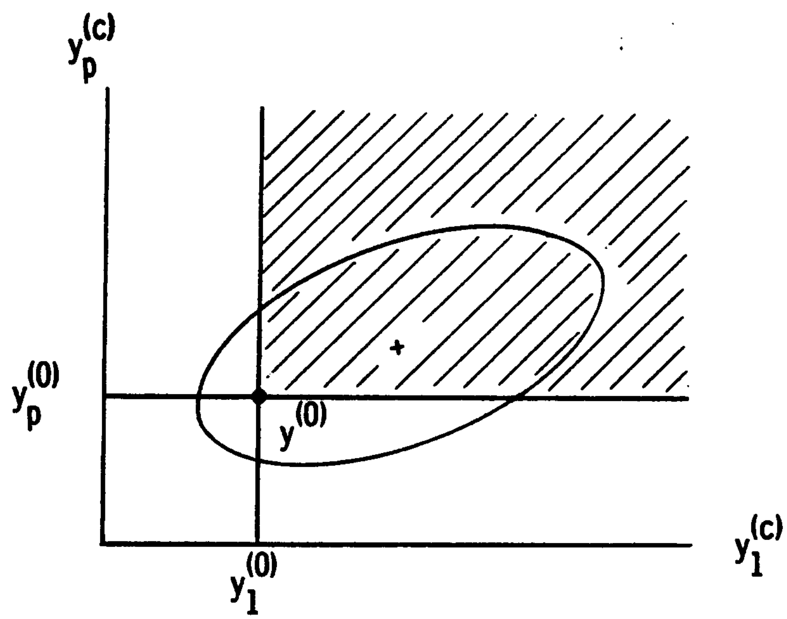
No	Response Matrix						Design Matrix					
	y_1	y_2	.	.	.	y_p	x_1	x_2	.	.	.	x_m
1	y_{11}	y_{21}	.	.	.	y_{p1}	x_{11}	x_{21}	.	.	.	x_{m1}
2	y_{12}	y_{22}	.	.	.	y_{p2}	x_{12}	x_{22}	.	.	.	x_{m2}
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
i	y_{1i}	y_{2i}	.	.	.	y_{pi}	x_{1i}	x_{2i}	.	.	.	x_{mi}
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
n	y_{1n}	y_{2n}	.	.	.	y_{pn}	x_{1n}	x_{2n}	.	.	.	x_{mn}

$\Omega : Y = X' B + U$,

$\{u_1', \ldots, u_n'\}$ are independently identically distributed with mean vector 0 and common $p \times p$ positive definite covariance matrix Σ .

Figure 3

PERFORMANCE RELIABILITY
MULTIVARIATE RESPONSE (VECTOR)



$$y^{(c)} \sim N[c' B, \Sigma],$$

$$R = P\{Y_1^{(c)} \geq y_1^{(0)}, \ldots, Y_p^{(c)} \geq y_p^{(0)}\}$$

$$= \int_{y_1^{(0)}}^{\infty} \cdots \int_{y_p^{(0)}}^{\infty} n(c' B, \Sigma)\, dy^{(c)}$$

Figure 4

APPENDIX A

Define $K = (y^{(0)} - c'\beta)/\sigma$ and $K(\hat{\beta},\hat{\sigma}) = (y^{(0)} - c'\hat{\beta})/\hat{\sigma}$. We first note that $\hat{\beta}$ and $\hat{\sigma}$ are independently distributed. Since $E\hat{\beta} = \beta$, we have that $E(y^{(0)} - c'\hat{\beta}) = (y^{(0)} - c'\beta)$. Also $v \equiv f\hat{\sigma}^2/\sigma^2$ has a χ^2_f distribution, $f = n - m$, so that

$$Ev^{-1/2} = \frac{\Gamma(\frac{f-1}{2})}{\sqrt{2}\ \Gamma(\frac{f}{2})} .$$

Hence,

$$E\,\frac{\sqrt{2}}{\sqrt{f}}\ \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\ \frac{1}{\hat{\sigma}} = \frac{1}{\sigma}$$

which proves that $\tilde{K}(\beta,\sigma) = \frac{\sqrt{2}}{\sqrt{f}}\ \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\ K(\hat{\beta},\hat{\sigma})$ is an unbiased estimator of $K(\beta,\sigma)$. By completeness, it then follows that $\tilde{K}$ is the unique such estimator, and hence is UMVU.

An alternative approach is also useful, namely, that $K(\hat{\beta},\hat{\sigma})/\|a\| \equiv t(f,\delta)$ has a non-central t-distribution with f degrees of freedom and non-centrality parameter

$$\delta = \frac{y^{(0)} - c'\beta}{\sigma\|a\|} = \frac{K(\beta,\sigma)}{\|a\|} ,$$

where $\|a\|^2 = c'(XX')^{-1}c$. To see this, we write

$$K(\hat{\beta},\hat{\sigma}) = \frac{\frac{y^{(0)} - c'\beta}{\sqrt{Var(c'\hat{\beta})}} - \frac{c'\hat{\beta} - c'\beta}{\sqrt{Var(c'\hat{\beta})}}}{\frac{\hat{\sigma}}{\sqrt{Var(c'\hat{\beta})}}} .$$

But $c'\hat{\beta} = c'(XX')^{-1}Xy$, and hence $var(c'\hat{\beta}) = \sigma^2 \ c'(XX')^{-1}c \equiv \sigma^2 \ \|a\|^2$. Thus,

$$\frac{K(\hat{\beta},\hat{\sigma})}{\|a\|} = \frac{\frac{y^{(0)} - c'\beta}{\sigma\|a\|} - \frac{c'\hat{\beta} - c'\beta}{\sigma\|a\|}}{\frac{\hat{\sigma}}{\sigma}} = t(f,\delta) .$$

By noting that $Et(f,\delta) = \delta \sqrt{f/2} \ \Gamma(\frac{f-1}{2})/\Gamma(\frac{f}{2})$, we can also obtain $E\ \tilde{K}(\beta,\sigma) = K(\beta,\sigma)$.

APPENDIX B

Since $R(\hat{\beta},\hat{\sigma}^2)$ is a function of the sample moments, it follows that $R(\hat{\beta},\hat{\sigma}^2)$ is asymptotically normal with mean $R(\beta,\sigma^2)$ and variance

$$\sum_{i,j} \left(\frac{\partial R}{\partial \beta_i}\bigg|_{\beta,\sigma^2}\right)\left(\frac{\partial R}{\partial \beta_j}\bigg|_{\beta,\sigma^2}\right) Cov(\hat{\beta}_i, \hat{\beta}_j) + \left(\frac{\partial R}{\partial \hat{\sigma}^2}\bigg|_{\beta,\sigma^2}\right) Var(\hat{\sigma}^2) .$$

The cross-product terms involving $\hat{\beta}_i$ and $\hat{\sigma}^2$ drop out because of the independence of $\hat{\beta}$ and $\hat{\sigma}^2$. From

$$R(b,s^2) = \int_{(y^{(0)} - c'b)/s}^{\infty} (2\pi)^{-1/2} \exp(-1/2 \ t^2) \, dt ,$$

we obtain

$$\frac{\partial R}{\partial b_i} = \frac{c_i}{\sqrt{2\pi}\, s} \exp[-1/2\frac{(y^{(0)}-c'b)^2}{s^2}], \quad \frac{\partial R}{\partial s^2} = \frac{(y^{(0)}-c'b)}{2s^2} \frac{1}{\sqrt{2\pi}\, s} \exp[-1/2\frac{(y^{(0)}-c'b)^2}{s^2}]$$

Also, $\mathrm{Cov}(\hat{\beta}_i,\hat{\beta}_j) = \sigma^2 a_{ij}$, where $A = (XX')^{-1}$, $\mathrm{Var}(\hat{\sigma}^2) = 2\sigma^4/f$, and hence the asymptotic variance is

$$V_\infty(\beta,\sigma^2) = \sigma^2[n(y^{(0)} \mid c'\beta;\ \sigma^2)]^2 \ [c'(XX')^{-1} c + \frac{(y^{(0)} - c'\beta)^2}{2\sigma^2 f}] \ .$$

But, $V_\infty(\hat{\beta},\hat{\sigma}^2)$ is a rational function of the sample moments, so that, by Slutsky's Theorem, $V_\infty(\hat{\beta},\hat{\sigma}^2)$ converges in probability to $V_\infty(\beta,\sigma^2)$, and hence

$$\frac{R(\hat{\beta},\hat{\sigma}^2) - R(\beta,\sigma^2)}{\sqrt{V_\infty(\hat{\beta},\hat{\sigma}^2)}} \to N(0,\ 1) \ .$$

APPENDIX C

The computation of the point and confidence interval estimates for the performance reliability of the warhead for the sample size $n = 8$ is described in this appendix. From the test data in Table I, we obtain the following results:

Point Estimation

$$\hat{\sigma}^2 = 2.958, \qquad c' = (1,\ 0,\ 0), \qquad f = n - m = 8 - 3 = 5$$

$$K(\hat{\beta},\hat{\sigma}) = \frac{y^{(0)}-c'\hat{\beta}}{\hat{\sigma}} = \frac{10 - 13.99}{1.720} = -2.320$$

$$\tilde{K}(\beta,\sigma) = K(\hat{\beta},\hat{\sigma})\sqrt{\frac{f}{2}}\ \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})} = -1.95$$

Substituting, these two estimates of $K(\beta,\sigma)$ in

$$R(\beta,\sigma^2) = \int_{K(\beta,\sigma)}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy$$

gives the two estimates of reliability $R(\hat{\beta},\hat{\sigma})$ and $\tilde{R}(\beta,\sigma)$, respectively. Thus,

$$R(\hat{\beta},\hat{\sigma}) = \int_{-2.320}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .9898\text{ , and}$$

$$\tilde{R}(\beta,\sigma) = \int_{-1.95}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .974 .$$

Confidence Intervals

$$\|a\|^2 = c'(XX')^{-1}c = \frac{1}{n} \text{ for } c' = (1, 0, 0) .$$

Following the notation of Resnikoff and Lieberman, a confidence interval may be obtained using the non-central t-tables in [6]. The percentage points of t are denoted by $x(f,\delta,\epsilon)$ where x is the value such that $P\{\frac{t}{\sqrt{f}} > x \mid f,\delta\} = \epsilon$.

$$x = \frac{K(\hat{\beta},\hat{\sigma})}{\sqrt{f}\ \|a\|} = \sqrt{\frac{n}{f}}\, K(\hat{\beta},\hat{\sigma}) = \sqrt{\frac{8}{5}}\,(-2.320) = -2.935 .$$

The one sided lower .95 confidence limit for R is obtained by finding the corresponding limit for $K(\beta,\sigma)$ because of the monotone relation between R and $K(\beta,\sigma)$. The $1-\alpha$ confidence limit for $K(\beta,\sigma)$ is obtained by solving

$$x(n-m, \delta_{1-\alpha}, 1-\alpha) = \frac{K(\beta,\sigma)}{\sqrt{f}\ \|a\|} ,$$

$$x(5, \delta_{.95}, .95) = -2.935 .$$

Making use of the relation $x(f,\delta,\epsilon) = -x(f,-\delta,1-\epsilon)$, we obtain

$$x(5, -\delta_{.95}, .05) = 2.935 .$$

From the Resnikoff-Lieberman table of percentage points of t, we obtain a non-centrality value $\delta = \sqrt{f+1}\ K_p = \sqrt{6}\ (1.107) = 2.712$ by interpolating on K_p. Since $\delta = \frac{K(\beta,\sigma)}{\|a\|}$, the .95 lower confidence limit for $K(\beta,\sigma)$ is

$$\underset{.95}{K(\beta,\sigma)} = \|a\|\ \delta_{.95} = \frac{\delta_{.95}}{\sqrt{n}} = \frac{-2.712}{\sqrt{8}} = -.959$$

Finally, the .95 confidence limit for R is

$$\underset{.95}{R(\beta,\sigma^2)} = \int_{\underset{.95}{K(\beta,\sigma)}}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .832 .$$

$\delta_{.95}$ may also be computed using the Johnson-Welch table IV and following the procedure on page 372 of [5].

This Page Intentionally
Left Blank

EVALUATION OF VARIOUS LABORATORY METHODS FOR DETERMINING RELIABILITY

A. Bulfinch
Reliability Concepts Section, Reliability Branch
Quality Assurance Division
Picatinny Arsenal

ABSTRACT. In preparation for evaluating several laboratory methods for determining reliability, an operational definition of reliability has been described in terms of stress and strength. The definition was verified by three independent methods of calculation.

Two classes of laboratory methods have been evaluated by means of Monte Carlo Sampling Procedures:

1. Testing-Without-Failure
2. Testing-With-Failure

The former type was found to be worthless or of limited value. The latter type was found to be capable of accurate and precise determination of reliability values at any level with small sample sizes.

I. INTRODUCTION. This investigation is concerned with the laboratory determination of ultimate* functional reliability with respect to single environments of components from systems to be used only once. To this end, Monte Carlo Experiments have been conducted to evaluate the precision, accuracy, and efficiency of two types of methods:

1. Methods which test-without-failure
2. Methods which test-to-failure.

In preparation for conducting these experiments, consideration was first given to establishing a valid procedure for calculating the true, ultimate[a] reliability of known conditions from the operational definition of reliability described below.

II. OPERATIONAL DEFINITION OF RELIABILITY. The theoretical definition of reliability, the probability of success under specified conditions, gives no clue as to how reliability can be experimentally determined. This

*Ultimate in this sense means maximum or unbiased value.

question, of course, must be answered before an experiment to measure reliability can be conducted. That is, the theoretical concept of reliability must be translated into an operational definition in terms of measurable properties and conditions. The following procedure is proposed as such a definition of the true, ultimate reliability of items to be used only once. The objective of the procedure is to measure the probability of failure--the complement of the probability of success.

The concept of reliability can be thought of in terms of stress and strength.* The relationship between these two elements can be represented graphically by two distributions on a single continuous scale: one (on the left) for the specified condition which will be called the applied stress curve and the other (on the right) for the property of an item which will be called the strength curve. The distance between the means of these distributions represents the margin of safety, or the extra strength built into the item during development. If the mean stress equals the mean strength, then the two curves are superimposed and the reliability is equal to 50%.

It is reasonable to assume that an item will not fail unless the applied stress equals or exceeds the item's strength. Referring to the two distributions described above, the probability of a strength value less than any particular stress value can be measured by the area under that portion of the strength curve which represents items with strengths less than that stress--that is, the area of the strength curve to the left of the stress ordinate. The probability of a particular stress value occurring can be measured by a small increment of area between two ordinates on the stress curve. In order to produce a failure in this manner, the higher stress value must occur simultaneously with the lower strength value. To determine the probability of such a combination of independent events occurring, the product of the probabilities of the separate events occurring must be obtained. There are many such mutually exclusive combinations possible. To obtain the total probability of failure, the sum of all the possible product must be calculated.

Defining the complement of reliability in this manner appears to be logically sound. The mathematics involved is based on well-known laws of probability. Therefore, the complement of the sum of the products of the probabilities of the stress equaling or exceeding the strength can be

---

*Strength here means the ability to withstand environmental stresses.

taken as the operational definition of the true, ultimate reliability of items used only once.

It would appear that combinations of conditions of this kind can represent, in an elementary way, the cause of failure when a component experiences an environment in actual use. A reliability obtained in this manner then has practical value and can be accepted as the solution to the basic problem of obtaining a valid, unbiased numerical value for reliability.

III. VERIFICATION OF THE OPERATIONAL DEFINITION.

A. Numerical Integration:

Having arrived at these conclusions the following combination of conditions was chosen for use:

Stress	U_1 = 10 (True average stress)
	S_1 = 5 (True std. dev. of stress)
Strength	U_2 = 20 (True average strength)
	S_2 = 5 (True std. dev. of strength)

Both distributions were assumed to be normal.

The reliability associated with this condition was calculated by the numerical integration procedure (calculating the sum of products) described above. The following results were obtained:

STRESS INCREMENT	RELIABILITY
1.0 Unit	.9362
0.5 Unit	.9264

B. Modified Normal Deviate:

A second method of determining the reliability value for the above condition which has its foundation for validity in statistical theory, is the following formula suggested by Dr. William S. Connor of the Research Triangle Institute:

$$Z = \frac{(X_1 - X_2) - (U_1 - U_2)}{\sqrt{S_1^2 + S_2^2}}$$

A failure can occur only when $X_1 \geq X_2$, where X_1 is any stress value and X_2 is any strength value. For this situation the formula becomes:

$$Z \geq \frac{U_2 - U_1}{\sqrt{S_1^2 + S_2^2}}$$

Substituting the above values assigned to U_1, U_2, S_1, and S_2:

$$Z \geq \frac{20 - 10}{\sqrt{25 + 25}} = \frac{10}{\sqrt{50}} = 1.414$$

Entering a table of areas under the standard normal curve, the following value can be obtained.

$$R \geq 0.9213$$

C. Monte Carlo Sampling:

A third method of determining the reliability value for the above condition is the Monte Carlo sampling of both distributions. This method is acceptable as valid on reasonable and logical grounds also. The stress-strength combinations that occur in actual use-conditions result from a random selection of such combinations. This is the basis of the Monte Carlo procedure described below which will be called the Flight-Condition Procedure.

Pairs of applied stress values and assumed strength values were chosen at random from their respective distributions by means of tables of random normal numbers from these distributions. The values in each pair were compared to determine whether the stress exceeded the strength. In those cases where this was true, the result of the observation was declared a failure.

The average reliability value obtained from 7,500 observations with the Flight-Condition Procedure was found to be R = 0.9221.

D. Summary of Results:

Method	Reliability
Numerical Integration	0.9264
Z-Formula	0.9213
Monte Carlo Sampling (Flight-Condition Simulation)	0.9221
Grand Average	0.9233

E. Conclusions of the Verification:

The close agreement of the values obtained by these three independent methods, each of which can be separately accepted as valid, is considered ample justification for concluding that:

1. The operational definition of reliability described above can be taken as valid.
2. The true-ultimate reliability for the condition described above is very close to the average of these three values: 0.9233.
3. The results obtained by the Z-Formula are sufficiently accurate to justify using this formula as the primary standard procedure for calculating the reliability of "one-shot" items under the above operational definition of reliability.

IV. EVALUATION OF LABORATORY METHODS.

A. Introduction:

In view of the above conclusions, the Z-Formula has been used to calculate the true-ultimate reliability values for the known stress-strength conditions used in the Monte Carlo Experiments that follow. These reliability values were used to compare with estimates obtained from the several laboratory methods being evaluated. The magnitude of the differences found between the true and estimated values were used to evaluate the accuracy of the results obtained from these methods.

The purpose of the Monte Carlo Experiments is to determine the accuracy, precision, and efficiency of various laboratory methods. This is done by Monte Carlo sampling of known distributions of stress and strength combinations.

B. Testing-Without-Failure Procedures:

1. Single Stress Level Method:

In practice, this method consists of applying the same environmental stress level to each test specimen comprising the sample and counting the number of failures obtained. The level of stress applied is usually one expected to be experienced in use. In the Monte Carlo Experiment, this method could produce a failure only when the value obtained from a table of random numbers was found to be equal to or less than the assumed stress value. The combination of conditions used for this experiment is:

Stress	Strength
$U_1 = 10$	$U_2 = 20$
$S_1 = 5$	$S_2 = 5$

In conducting this Monte Carlo Experiment, the following assumptions were made:

(1) The stress being applied was 10 units.

(2) The tabled values of random numbers used contained a decimal point to the left of the first digit and that, as such, these decimal values represent the probability associated with the strength values used.

(3) A failure was obtained only when a tabled value was found to be less than 0.0228 - the probability of a strength value equal to or less than 10 units.

Sample sizes of 22, 35, 45, and 230, were used. One hundred samples of each size were taken. The purpose here is to determine the effect of increasing the sample size in addition to the purpose stated above.

a. Results:

TABLE I

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 22

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	73
1	.045		23
2	.091		4

For : 0/20 : 90%(1 - sided) upper limit for defects = .099.

The true probability of failure is included in the 90% one-sided upper confidence limit.

TABLE II

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 35

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	57
1	.029		35
2	.057		7
3	.086		1

For 0/35 : 90% (1 - sided) upper limit for defects = .063

95% (1 - sided) upper limit for defects = .082

The true value is included in only the 95% upper limit.

TABLE III

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 45

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	50
1	.022		36
2	.044		14

For 0/45: 90% (1 - sided) upper limit for defects = .049

95% (1 - sided) upper limit for defects = .064

99% (1 - sided) upper limit for defects = .097

The true value is included in only the 99% upper limit.

TABLE IV

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 230

Failures No.	Proportion		Frequency, Per cent
0	0.000		0
1	.004		4
2	.008	80%(2-sided) lower limit	10
3	.013		7
4	.017	Mode	21
5	.022		17
6	.026		14
7	.030		11
8	.035	80%(2-sided) upper limit	7
9	.039		4
10	.043		4
11	.048		0
12	.052		0
13	.056		1

4/230: 99.5% (1 - sided) upper limit for defects = .053
The true value is not included in even the 99.5% upper limit.

TABLE V

SUMMARY

SINGLE STRESS LEVEL METHOD

True Probability of Failure: P = 0.0787

True Reliability: R = 0.9213

EFFECT OF SAMPLE SIZE

Sample Size	Average R	Average P
22	.987	.013
35	.985	.015
45	.986	.014
230	.977	.023

b. Conclusions:

The following conclusions can be drawn from the above data about the method which applies a single stress level under the use conditions:

(1) This method produces biased results.

(2) Increasing the sample size did not remove the bias or error.

(3) The confidence interval is worthless for locating the true value.

(4) The ultimate value of the failure rate being measured by this method is 0.0228, the probability of a strength value equal to or less than 10 units - not the true probability of failure as previously defined.

(5) The possibility of a stress value exceeding a strength value greater than 10 units is completely ignored by this method,

(6) This method is worthless to determine reliability as defined above.

2. Multiple-Stress Level Method:

The flight-condition procedure used earlier as one of the intuitively acceptable basic procedures for determining true reliability, is one which also readily lends itself to laboratory use with small samples. For this purpose, a different randomly selected stress is used for each test specimen.

a. Results:

Using the Flight-Condition Procedure in the Monte Carlo Experiment with sample size of 50, the following results were obtained:

TABLE VI

MULTIPLE - STRESS LEVEL METHOD

(Low Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size = 50

DISTRIBUTION OF FAILURES FOR SAMPLES OF 50

Number of Failures	Frequency	
0	2	
1	6	
2	28	
3	27	
4	39	Mode
5	25	
6	11	
7	5	
8	4	
9	3	
Total Number of Sample Results:	150	

Average Reliability = 0.9224

Standard Deviation = 0.0376

TABLE VII

MULTIPLE - STRESS LEVEL METHOD

(High Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9977

Sample Size = 50

DISTRIBUTION OF FAILURES FOR SAMPLES OF 50

Number of Failures	Frequency	
0	90	Mode
1	9	
2	1	
3	0	
Total Number of Sample Results	100	

Average Reliability = 0.9978

Standard Deviation = 0.0069

b. Conclusions:

Again the Flight-Condition Procedure produced an accurate, unbiased average value. However, in the case of high reliability, this procedure produced no failures in 90% of the samples of 50. In addition, no results could be obtained between 1.00 and .98. This is a distinct disadvantage. In practice only one small sample is usually taken. If no failures are obtained, the result is not only biased but it is worthless for mathematical manipulations of any kind. From this it is concluded that testing under the expected stress conditions leaves something to be desired in determining high reliabilities. This type of method is too insensitive to changes in high reliabilities to be of use with small samples.

V. TESTING-TO-FAILURE PROCEDURES.

A. Introduction:

An alternative approach for determining reliability is "Testing-to-Failure". Instead of applying the stress or stresses expected in use, systemstically increase the level of severity of the stress until failure occurs. Methods for this purpose are available which will generate the ultimate strength distribution. From the results of this type of testing, the average and standard deviation of the ultimate strength can be determined. By means of independent experiments or prior knowledge, the average and standard deviation of the expected stress distribution can be obtained.

1. T-Formula:

With these two sets of data available, reliability, as previously defined, can be estimated from the Z-Formula by replacing the population means and variances with their sample estimates:

$$T \geq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 + s_2^2}}$$

where:

$\bar{X}_2$ = Observed Average Ultimate Strength

$\bar{X}_1$ = Observed Average Expected Stress

s_1^2 = Observed Variance of the Expected Stress

s_2^2 = Observed Variance of the Ultimate Strength.

Using tables of areas under the normal curve to obtain reliability values from the T-Formula for samples as small as 20, causes an error in the result. However, this error is small -- less than one per cent.

2. Need for Indirect Methods:

Testing-to-Failure Procedures can be easily performed when the measure of a simple property, such as tensile strength, is required. This is the ideal situation in which:

a. The occurrence of a failure can be detected by visual inspection.

b. The magnitude of the applied stress, at the point of failure, is directly observable.

The results obtained in a test of this kind, permit the calculation of the average and variance of the ultimate strength directly from the observed data. These values can then be used in the T-Formula to calculate the reliability with respect to tensile load.

To measure more complicated properties, such as the integrity of a moisture seal after vibration, indirect methods must be used to obtain the average and variance of the ultimate strength. In this case:

a. The occurrence of a failure cannot be detected by visual inspection. A failure must be determined by means of the results obtained from conducting an appropriate test.

b. The magnitude of the stress at the point of failure is not directly observable. The stress at the point of failure must be calculated.

The observed data in this kind of test is simply success or failure. As the stress is applied at higher and higher levels, higher and higher proportions of the test specimen will fail. If the stress is increased step-wise, and the proportion of failures determined at several points between 0 and 100% failure, the familiar sigmoid-curve of a cumulative frequency distribution can be obtained from these results.

The average and variance of the ultimate strength can be determined from this curve. The average value is the stress level causing 50% failures. The variance is the square of the reciprocal of the slope of this curve at the 50% point.

3. Characteristics of Available Methods:

The class of methods that can generate curves of this type are called Tests-of-Increased Severity. Some of these methods are:

a. The Run-Down
b. The Two-Stimuli
c. The Up-and-Down

The Run-Down Method can be used to determine the shape of the sigmoid-curve. But, it requires the largest number of test specimens.

The Two-Stimuli Method requires the assumption of normality. It is:

a. An abreviated Run-Down Method

b. Highly efficient when the above assumption is valid, since it takes a minimum number of test specimens

c. Easy to conduct

d. Easy to calculate the average and variance.

The Up-And-Down Method is highly efficient for determining the 50% point. However, it gives a poor estimate of the variance without large sample sizes. In addition, this method is difficult to conduct and requires lengthy grouped data calculations to obtain the average and variance.

4. Two-Stimuli vs Up-And-Down Method:

In view of these characteristics and the following data, the Two-Stimuli Method was chosen for conducting most of the initial work on this class of methods.

a. Results:

TABLE VIII

Test-to-Failure Procedures

Conditions Used:	STRESS	STRENGTH	
	$U_1 = 10$	$U_2 = 20$	(True averages)
	$S_1 = 5$	$S_2 = 5$	(True standard deviations)

True Reliability = .9213

METHOD	TOTAL NO. OF TRIALS	AVG.	STANDARD DEVIATION	RELIABILITY
Up-and-Down	2700	19.8	4.88	.9197
Two-Stimuli	1600	20.0	5.03	.9207

b. Conclusions:

Any observed differences in the averages or standard deviations are highly significant for these large sample sizes. Therefore, basing a choice on the absolute magnitude of the observed differences, it can be said that the Two-Stimuli Method is slightly more accurate and precise than the Up-and-Down Method.

B. Two-Stimuli Method:

1. Description:

The Two-Stimuli Method was conducted as follows:

Starting at an assumed stress value of 10 units, the first tabled value (converted to a decimal by placing a decimal point to the left of the first digit) was compared to .0228. If the tabled value was found

to be greater than .0228, the result was declared a success. Then a second tabled value (as a decimal) was chosen and compared to .0668, the probability value associated with a stress of 12.5 units. If the tabled value was greater than .0668, the result was again declared a success. Then a third tabled value was compared with .1587, the probability associated with a stress of 15 units. This process was continued using increments of stress equal to $s_1/2$, until the tabled value (as a decimal) was found to be equal to or less than the probability value associated with the assumed stress. When such a comparison was found, the result was declared a failure. Only half of the cases of exact equality were declared failures. At this level of stress, 19 additional observations were made by the comparison process just described. The proportion of failures at this level of stress was calculated and recorded. If this proportion was less than 50%, then the stress level was increased by two or three increments. If the proportion of failures was greater than 50%, then the stress level was decreased by two or three increments. A total of 20 observations was made, as before, at this latter stress level; the proportion of failures calculated and the result recorded. Only proportions greater than 0 and less than 100%, and which differ by 20% or more, are useful for this purpose. The average and standard deviation of the strength curve was calculated from these two failure proportions by means of the equation for the standard normal cumulative frequency curve.

2. Results:

The following results were obtained using the Two-Stimuli Method:

TABLE IX

TWO-STIMULI METHOD
(Low Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size = 45

DISTRIBUTION OF RESULTS FROM SAMPLES OF 45

Reliability Cell - Width	Frequency	
.990-.970		
.969-.950	17	
.949-.930	42	Mode
.929-.910	25	
.909-.890	32	
.889-.870	14	
.869-.850	12	
.849-.830	3	
.829-.810	1	
.809-.790	3	
.789-.770	1	
Total Number Of Sample Results	150	

Average Reliability = 0.9136

Standard Deviation = 0.0369

TABLE X

TWO-STIMULI METHOD

(High Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = .9977 Sample Size = 45

DISTRIBUTION OF RESULTS FROM SAMPLES OF 45

Reliability Cell-Width	Frequency
.9999 - .9998	2
.9997 - .9990	41
.998 - .995	74 (Mode)
.994 - .991	20
.990 - .987	6
.986 - .983	2
.982 - .979	0
.978 - .975	1
.974 - .971	0
.970 - .967	2
.966 - .963	1
.962 - .959	0
.958 - .955	0
.954 - .950	1
TOTAL NUMBER OF SAMPLE RESULTS	150

Average Reliability = 0.9952

Standard Deviation = 0.0064

3. Conclusions:

It is clear from the above results that estimates of reliabilities at any level can be obtained with acceptable accuracy and precision from small samples when using the Two-Stimuli Method. This method is sensitive to changes that actually take place in either low or high reliability conditions. In principle then, this type of method is the solution to the problem of measuring high reliabilities with small sample sizes.

VI. SUMMARY:

A. Comparison of Testing-With-Failure with the best of the Testing-Without-Failure Methods.

TABLE XI

1. LOW RELIABILITY

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size: Two-Stimuli Method, N = 45
Flight Condition Method, N = 50

DISTRIBUTION RELIABILITY RESULTS

Reliability Cell Midpoints	Two-Stimuli	Flight Condition
1.00	0	2
.98	0	6
.96	17	28
.94	42 (Mode)	27
.92	25	39 (Mode)
.90	32	25
.88	14	11
.86	12	5
.84	3	4
.82	1	3
.80	3	0
.78	1	0
Average Reliability =	.9136	.9224
Standard Deviation =	.0369	.0366

TABLE XII

2. HIGH RELIABILITY

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9977

Sample Size: Two-Stimuli Method, H = 45
Flight-Condition Method, N = 50

DISTRIBUTION OF RELIABILITY RESULTS

Reliability Cell-Width	Flight-Condition Method	Two-Stimuli Method
1.0000 - .9998	90 (Mode)	2
.9997 - .9990	0	41
.998 - .995	0	74 (Mode)
.994 - .991	0	20
.990 - .987	0	6
.986 - .983	0	2
.982 - .979	9	0
.978 - .975	0	1
.974 - .951	0	0
.970 - .967	0	2
.966 - .963	0	1
.962 - .959	1	0
.958 - .955	0	0
.954 - .951	0	1
TOTAL NUMBER OF SAMPLE RESULTS:	100	150
Average Reliability	0.9978	0.9952
Standard Deviation	0.0069	0.0064

B. Conclusions:

There is little to choose between these two methods for reliability values equal to or less than .9213. However, for reliability values as high as .9977, the Two-Stimuli Method appears to be superior for practical use when numerical values for reliability are required.

It can be seen from Table X that the Two-Stimuli Method never produces a result of 1.000 for reliability. This is in contrast to the Flight-Condition Procedure which produced a reliability value of 1.000 in 90% of samples of 50 (Table VII). Seventy-eight percent of the Two-Stimuli Method results are within $\pm$ 0.002 of the true value for the high reliability condition. In the Flight-Condition Procedure, no result can be obtained between 1.000 and .9800 when a sample size of 50 is used.

VII. CONCLUSIONS:

1. For practical engineering purposes, reliability can be defined in terms of stress and strength. From this, three simple conclusions follow:

   a. Reliability is created by the strength built into an item.

   b. An item cannot fail until the stress equals or exceeds the strength.

   c. To increase the reliability above 50%, the strength must exceed the stress.

2. Reliability is a relative property that depends upon the environmental stresses to be experienced (in use). To state a numerical value for this property requires a careful definition of the expected stresses.

3. The magnitude of a reliability value does not depend upon the number of items tested. Increasing the sample size can only improve the precision with which the reliability value is known. This improvement can be obtained only if the method of testing produces an unbiased estimate of reliability.

4. Applying only a single level of the expected stress to all of the test specimens cannot determine reliability at any level with an acceptable degree of accuracy and precision.

5. Applying a different randomly chosen stress level, from the expected stress distribution, to each test specimen used, produces unbiased results and is only useful for determining low reliability values.

This procedure cannot determine high reliabilities with small sample sizes with an acceptable degree of precision.

6. The Two-Stimuli Method can determine reliability at any level with small sample sizes with an acceptable degree of accuracy and precision.

COMPUTER SIMULATIONS IN RELIABILITY

Jasper Dowling
U. S. Army Munitions Command, Picatinny Arsenal

ABSTRACT. The troublesome problems of calculating a realistic lower confidence limit for systems reliability from component results and writing algebraic probability expressions for complex systems have been investigated. Practical Monte Carlo procedures for routine use of high speed computers are described. Iterative procedures are explained which can:

a. Save time in calculating the lower confidence limits of complex systems.
b. Obtain point estimates of complex systems from Boolean expressions when the writing of algebraic equations is too difficult.

INTRODUCTION. This paper deals with two of the problem areas encountered in obtaining estimates of the reliability of a weapon system from component test data. The accuracy of the reliability estimate of a system is a function of the accuracies of the component reliability estimates. The calculation of the lower bound of the system estimate, or lower confidence limit, is one of the problem areas and is the first of the two discussed. The remainder of this paper explains a computer simulation technique for evaluating system Boolean expressions where algebraic probability expressions are not feasible.

CONCLUSION.

1. It is concluded that the Monte Carlo technique is a valid and practical method of obtaining the lower confidence limit of a system. A 90 % lower confidence limit of 0.88 obtained with the Monte Carlo method corresponds with that obtained by Garner and Vail (Reference 2) using a three component system in a series configuration with component reliabilities of 0.96, 0.97, and 0.99. Their 95 % lower confidence limit of 0.88 using a different technique corresponded with the lower limit of 0.88 obtained with the Monte Carlo method.

2. A five component system of known system reliability was used as a standard in the second part of this report. The reliability of this known system was 0.98. By using the formula

$$\bar{P} \pm 2\sqrt{\frac{pq}{n}}$$

it was determined with 95 % assurance that a Monte Carlo sample size of 78,400 would be needed to obtain values of 0.98 ± .0005. The system was run and resultant values of 0.9796 and 0.9799 were obtained. Although this is only an example it is concluded that the application of the simulation is valid, and hence the procedure is a practical, useful one.

MAIN DISCUSSION.

1. To begin the discussion of the lower confidence limit let us assume a three component system in a series configuration. See Figure 1[1]. These components are assumed to be independent. This system reliability is 0.96 x 0.97 x 0.99 = 0.92 (to two decimal places). In this example each success ratio and hence, each binomial probability distribution is depicted for a sample size of 100. By using the binomial probability law we can compute a 90 % lower confidence limit for each of the components, as 0.92, 0.94, and 0.96 respectively. Their product 0.92 x 0.94 x 0.96 = 0.83. This is not a 90 % confidence limit of the system.

2. To achieve the desired system distribution a Monte Carlo technique can be used to perform the product of the three component distributions. The probability of choosing a particular value from the component distribution will have to be equal to the distribution's probability (ordinate). To obtain this with a uniform random number generator, the three component binomial distributions are put into cumulative distributions. See Figure 2. A random number x, $0 \leq x \leq 1$ is generated. If this random number is less than the first value on the cumulative distribution (lowest ordinate) the value of the abscissa at this ordinate is assigned to the value of this component. If the random number is greater than the first ordinate, it is compared to the second ordinate. This continues until an ordinate is greater than the random number. The corresponding abscissa is assigned the value of the component. This is done for each of the three components and the three assigned values are substituted into the system equation and the equation solved. This is one point of the system distribution. As an example, assume the three random numbers .3321645, .2168429O, and .9316420O are chosen. The corresponding reliabilities would be 0.95, 0.96, 1.0. The point on the system distribution would be the product of these three or 0.912O. This procedure is repeated many times to form the system distribution.

[1] All figures are contained in the appendix.

3. Figure 3 is the system distribution of Figure 1. This distribution is based on 5,000 points (repetitions) with a mean of 0.92, a standard deviation of 0.026 and a 90 % lower confidence limit of 0.88. This lower confidence limit can be obtained by assuming normality and calculating the limit or by counting the lowest 500 points (10 % of the total) of the system frequency distribution .

4. Figure 4 represents a 22 component system. Each component is represented by its binomial probability distribution based on a sample size of 100. It should be noted that although there are a number of similar components to be assigned the same probability, the algebraic equation must allow for 22 independent components. Assigning similar components the same Monte Carlo or simulated value will cause both higher and lower system values and hence an excessively high variance and a wider frequency distribution. For accuracy in counting the lowest cells for determining a counted confidence limit, the system frequency distribution is collected in small cell intervals and grouped after counting.

5. Figure 5 is the system distribution of Figure 4 based on 5,000 points. The mean value of this 22 component system is 0.98 and the lower 90 % confidence limit is 0.97. It is interesting to point out that this distribution looses its symmetry as the sample size is decreased. The left hand tail becomes quite long and tapered. It should also be noted that although this distribution is essentially binomial, the sample size of the distribution becomes obscured in the formation of the distribution.

6. The second phase of this paper will be devoted to procedures for the evaluation of a system where the algebraic probability expression is not available. While dependency of components is a contributor in making the algebraic probability expression difficult, a system algebraic probability expression can become "not feasible" even when all components maintain their independence. Examples of situations that add to the complexity of a system probability expression are:

a. Mechanical and electrical couplings

b. One part of a system functions only if a second part of the system fails

c. Multi-option channels.

7. Consider Figure 6. Clearly this is not a simple series-parallel diagram. There are 16 paths of success through this diagram, $A_1B_1C_1D_1E_1F_1$ etc. To write this as a system of 16 in parallel would require a considerable amount of algebra to account for repeating components in more than one of the 16 success paths. One approach to a solution to this problem is to redraw the diagram in a simple series-parallel configuration. See Figure 8. The trouble with this diagram is that there are twenty-four components represented where actually there are only nineteen physical components. Thus, the twenty-four components are not independent and in order to represent the system in a simple series-parallel configuration, it is necessary to draw the same component more than once. This procedure perhaps lessens the algebra required to write an algebraic probability expression, however, it would be preferable to consider a technique that doesn't require indepencence or an algebraic equation.

8. The basic notation and concept of Boolean algebra should be mentioned at this time. A plus sign is used to mean "or", and a dot is used to mean "and". The following expressions are thus introduced:

$$1 + 1 = 1 \qquad 1 \cdot 1 = 1$$

$$1 + O = 1 \qquad 1 \cdot O = O$$

$$O + 1 = 1 \qquad O \cdot 1 = O$$

$$O + O = O \qquad O \cdot O = O$$

9. For purposes of illustration the example used earlier will be used again. Consider Figure 7. The procedure is as follows: Generate a random number (RN) from a uniform distribution:

If RN $\leq$ reliability of component A, assign A = 1

If RN $\geq$ reliability of component A, assign A = O.

Components B and C are treated similarly. Now, the probability expression in Boolean notation is $P = A \cdot B \cdot C$ which reads P = A and B and C. In order for P to be a "1" all three components must be "1".

The probability of assigning a "1" to a component is the success ratio of the component. Hence, the probability of assigning all three components a "1" is the system's probability of success.

10. From Figure 8 the following expressions for different groups of components (X) can be obtained.

$$X_1 = A_1 \cdot B_1 + A_2 \cdot B_2$$

$$X_2 = C_1 \cdot (D_1 \cdot E_1 + D_3 \cdot E_2)$$

$$X_3 = C_2 \cdot (D_2 \cdot E_1 + D_4 \cdot E_2)$$

$$X_4 = X_1 \cdot (X_2 + X_3) \cdot F_1$$

$$X_5 = A_1 \cdot B_3 + A_2 \cdot B_4$$

$$X_6 = (C_3 + C_4) \cdot (F_2 + E_3 \cdot F_1)$$

$$PROB = X_4 + (X_5 \cdot X_6)$$

11. To solve this final expression a random number is generated and compared with component one, A_1. A_1 is assigned a "O" or "1". This process continues until all nineteen (not twenty-four) components have been assigned a "O" or "L". IMPORTANT: when A_1, A_2, E_1, E_2, and F_1 are assigned a value they are assigned the same value everywhere they appear. Upon solving, X_1 will be a "1" or "O". The probability of success of the system can be represented by

$P = \frac{1}{n} \sum_{i=1}^{n} PROB_i$ where n is the number of times the system is simulated.

References

1. Orkand, Donald S., "A Monte Carlo Method for Determining Lower Confidence Limits for System Reliability on the Basis of Sample Component Data", Picatinny Arsenal, Dover, New Jersey, June 1960.

2. Garner and Vail, "Confidence Limits for System Reliability", Systems Design, September-October 1961.

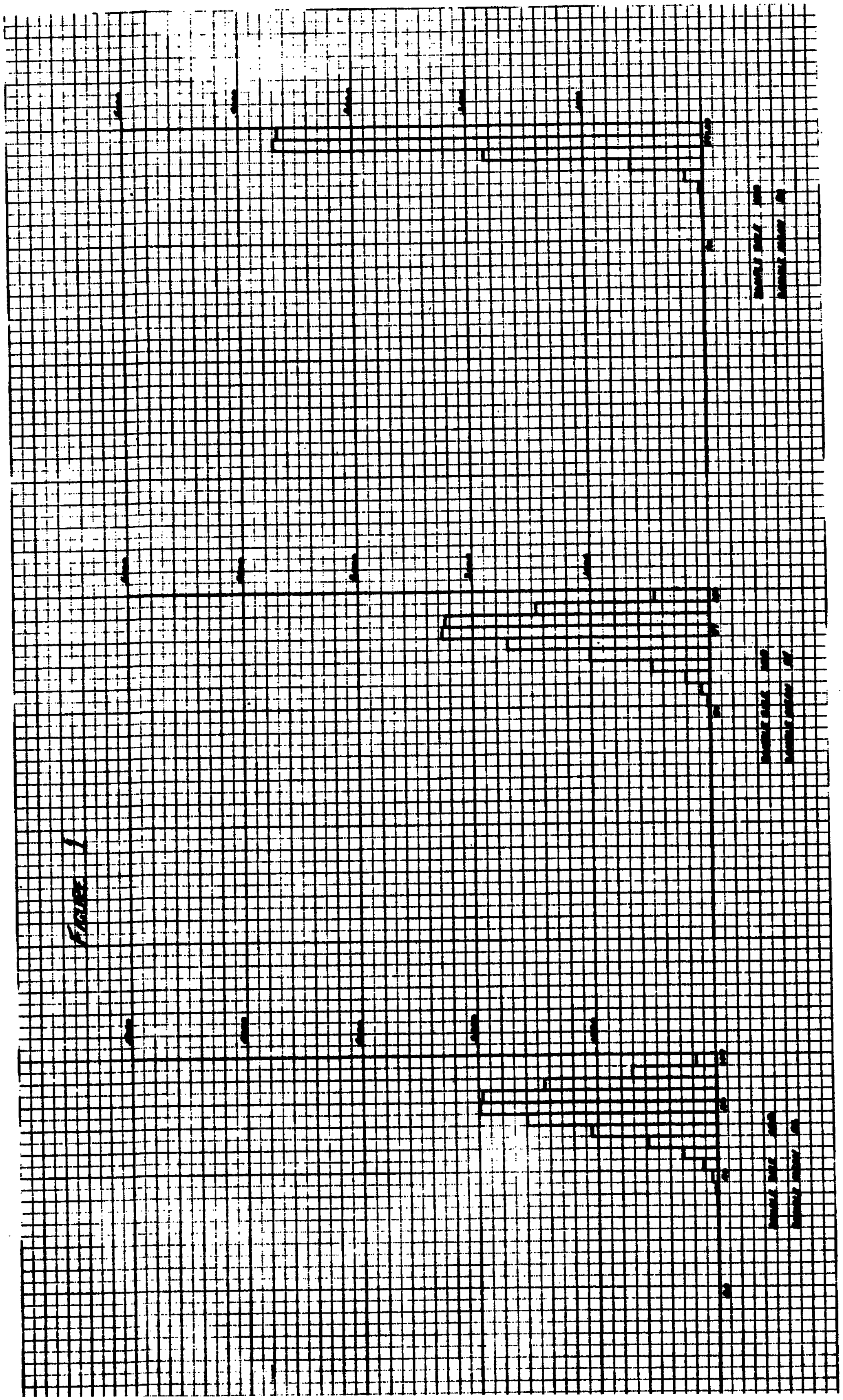


FIGURE 1

This Page Intentionally
Left Blank

536

FIGURE 2

This Page Intentionally Left Blank

538

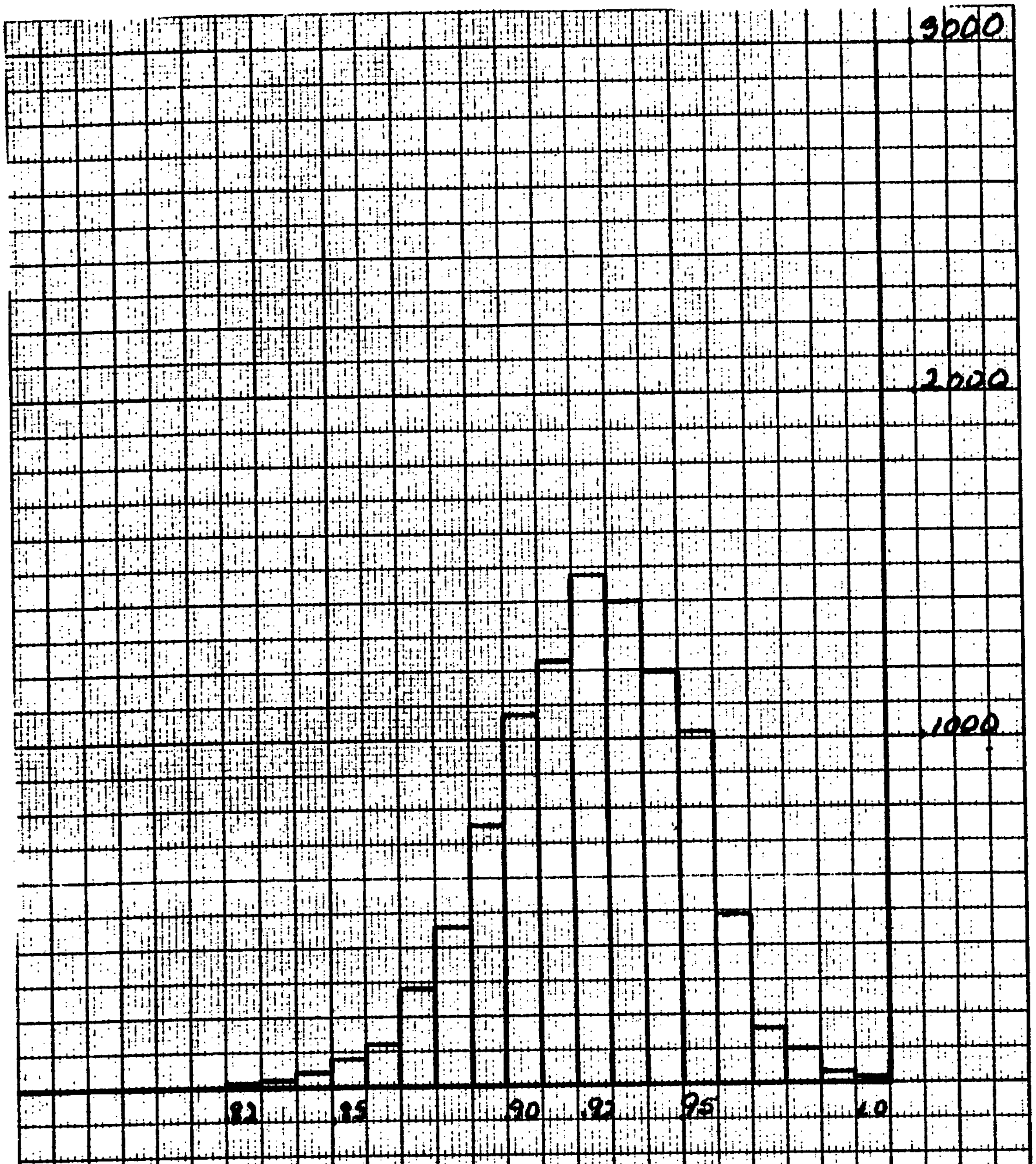
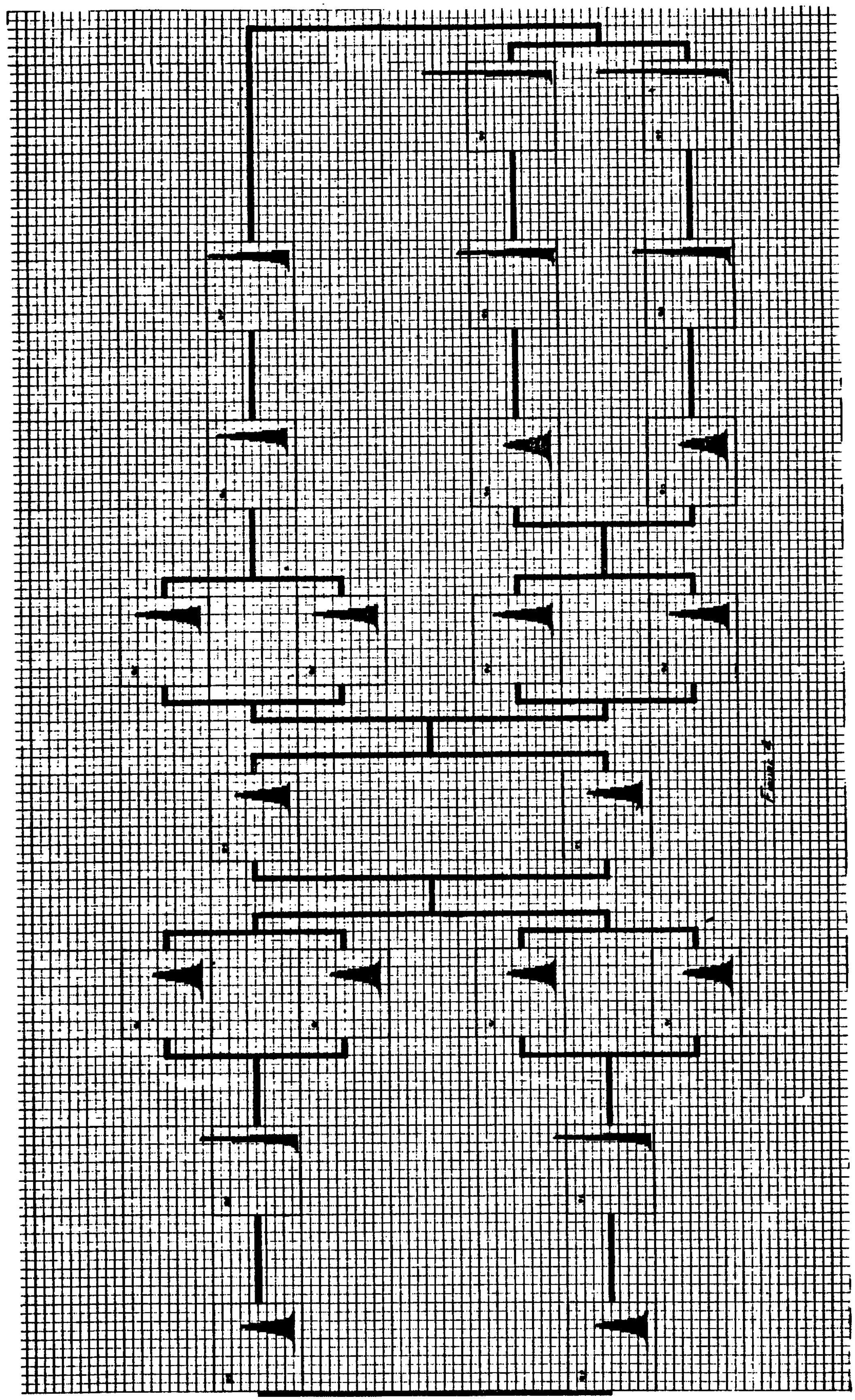


FIGURE 9

This Page Intentionally
Left Blank

540



This Page Intentionally
Left Blank

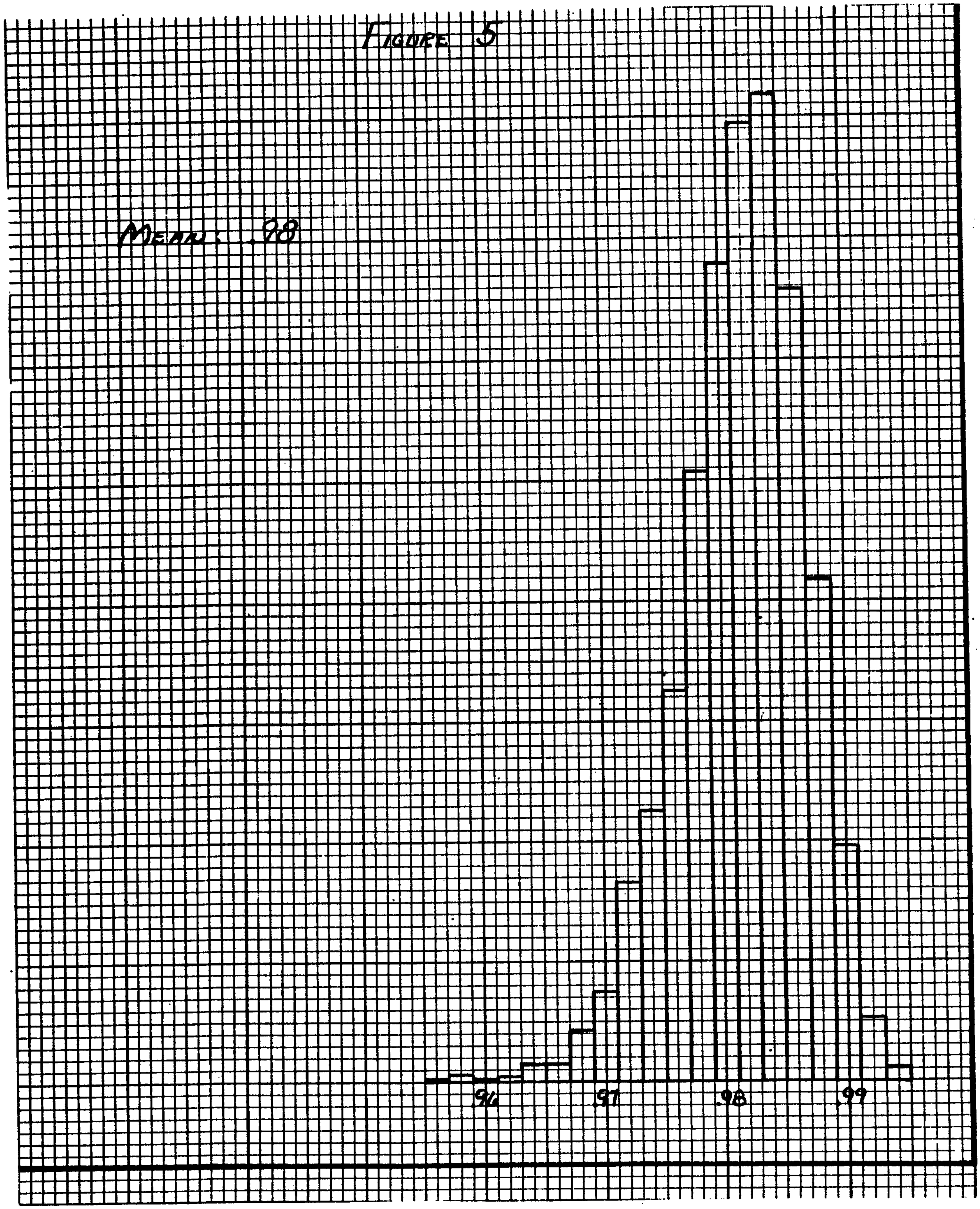


FIGURE 5
MEAN: .98
.96
.97
.98
.99

This Page Intentionally
Left Blank

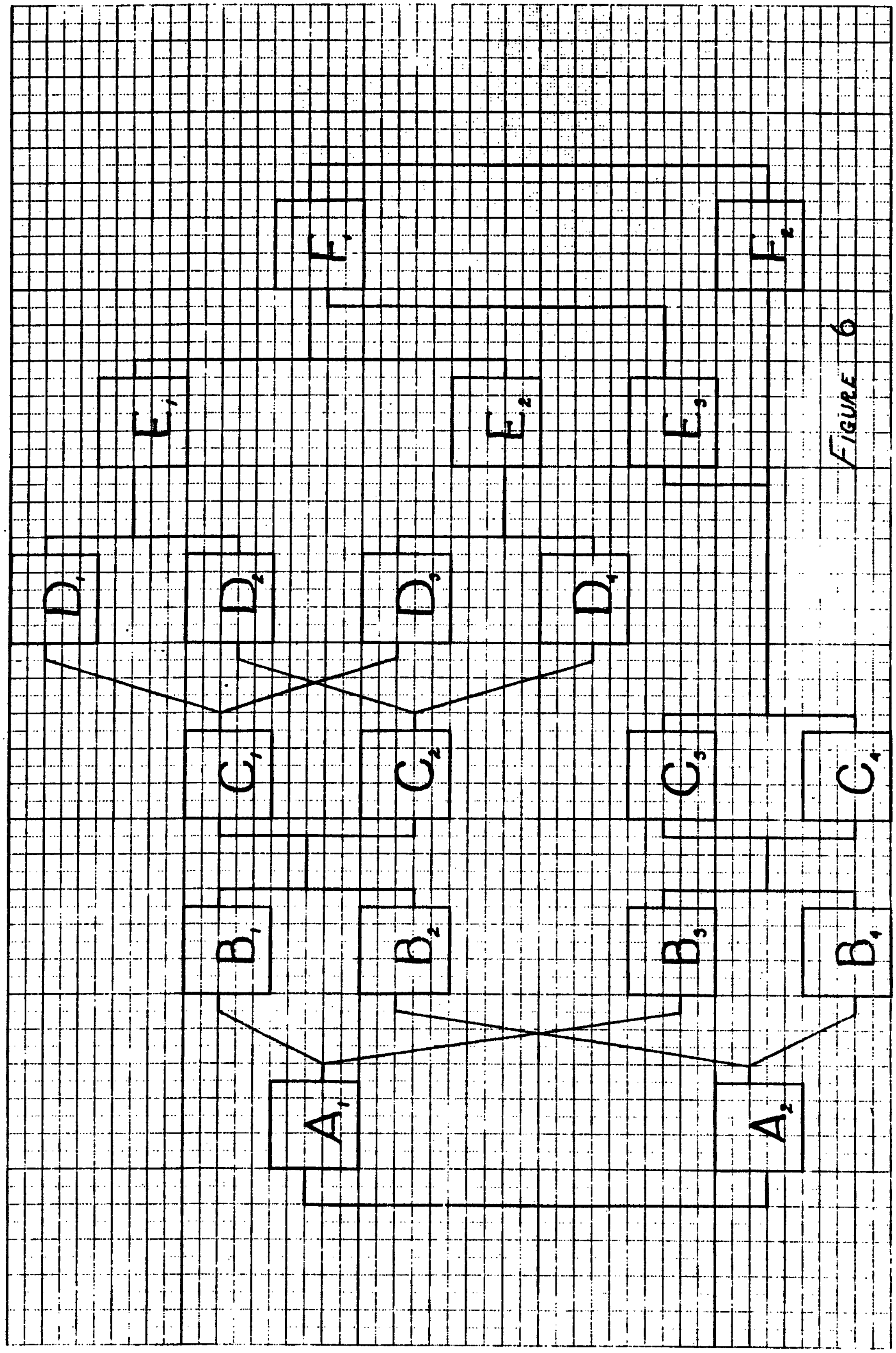


Figure 6

This Page Intentionally
Left Blank

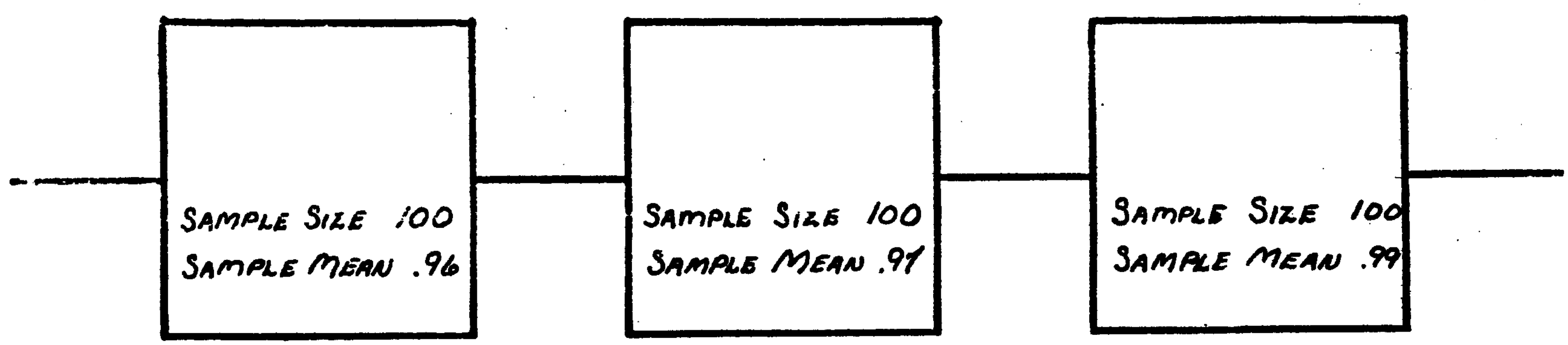

SAMPLE SIZE 100
SAMPLE MEAN .96
SAMPLE SIZE 100
SAMPLE MEAN .97
SAMPLE SIZE 100
SAMPLE MEAN .99

FIGURE 7

This Page Intentionally
Left Blank

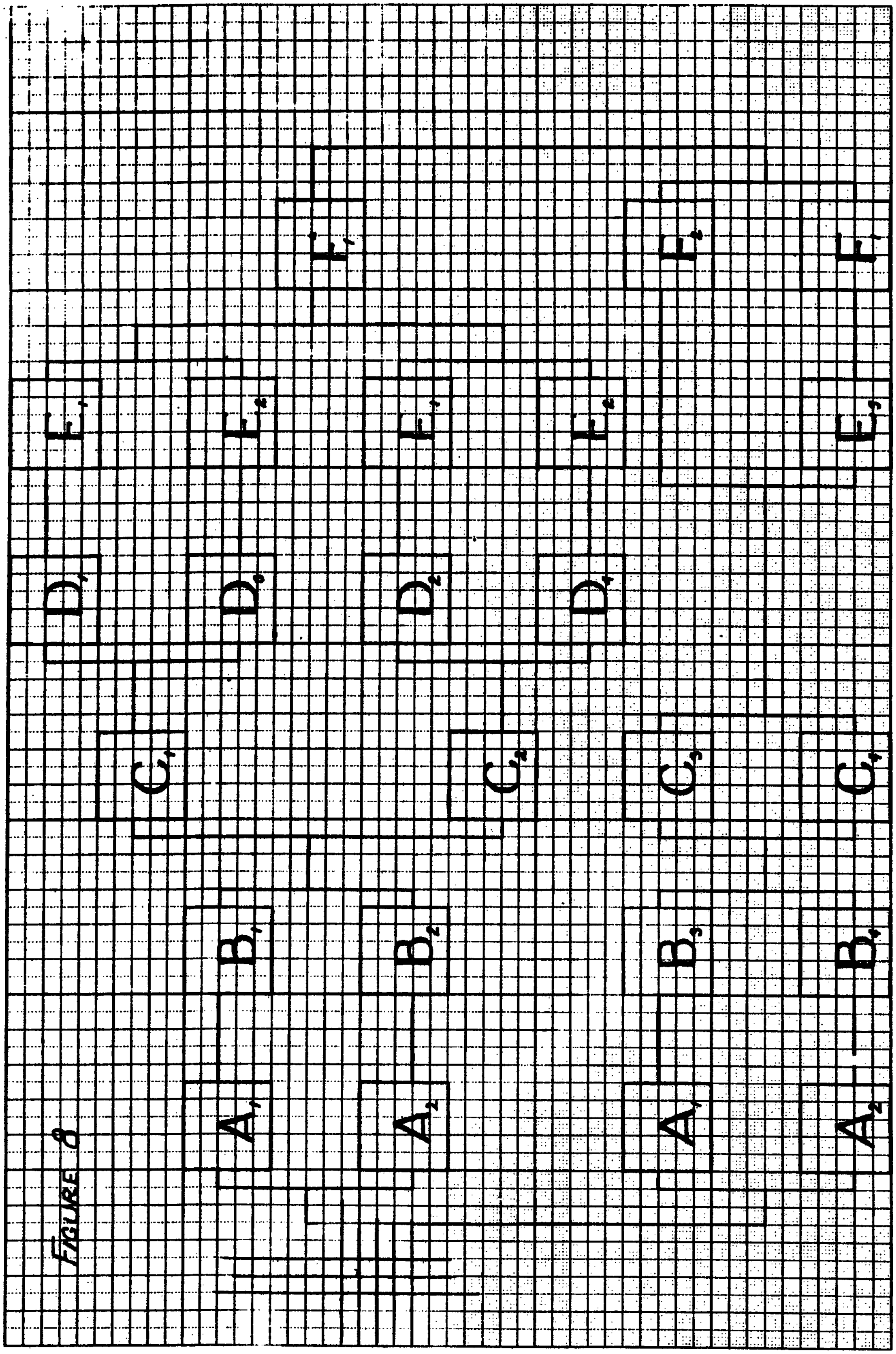


FIGURE 8

This Page Intentionally
Left Blank

BIOASSAY: THE QUANTAL RESPONSE ASSAY

H. C. Batson
University of Illinois College of Medicine
Chicago

I. INTRODUCTION. In many instances of interest in medical and biological research, the properties, activity or potency of certain substances cannot be measured directly by common *in vitro* chemical or physical methods, but can be measured (quantitated) only in terms of some effect they evoke in a living test subject, - animal, plant or microorganism.

Substances in this category include many hormones, vitamins, pharmacologically and toxicologically active substances, antibiotics, and immunologically active substances, - vaccines, toxins, toxoids, antisera, allergens, etc. Measurement or quantitative assessment of the activity of such substances constitutes the subject matter of biological assay.

Design of bioassay experiments and statistical analysis of the resultant data involve mainly an extension of principles and procedures readily available in standard references on experimental design and statistical analysis with major emphasis on regression analysis and analysis of variance with or without transformation of the data originally recorded in conventional units.

II. TYPES OF BIOASSAYS

1. On a basis of intent: On the basis of intent, bioassays can be classified in one of two main groups, - absolute or comparative.

Absolute assays: Absolute assays involve an attempt to obtain some quantitative measurement that can be expressed in absolute terms, such as a Minimal Lethal Dose (MLD) or Median Effective Dose (ED_{50}, LD_{50}, etc.). Such attempts are based on the assumption or belief that some such absolute value exists and that universally it can be determined with adequate precision. However, the absolute potency of substance X for "the cat" typically depends on just which cat is used and, unfortunately, cats invariably do differ. Laudable though the goals and objectives may be, absolute assays of biologically active substances, with

few (if any) exceptions, have little useful quantitative meaning.

Comparative assays: Although absolute assays seldom if ever yield adequately reproducable results, it generally is possible to achieve experimental quantitation of many biologically active substances through assessment of the substance of interest (unknown) in direct comparison with a reference substance (standard) qualitatively identical or, at least, similar in terms of the response evoked in the test subject of choice. While the absolute potency of either may never be known, the comparative or relative activity of the two may be assessed and the biological activity of the unknown expressed in relation to that of the standard in terms of relative potency, - whether expressed in proportions, percentages or in arbitrarily defined units. By using a common reference or standard substance, various investigators may obtain quantitative results with a degree of comparability adequate for their needs. Such relative potency estimates are subject to uncertainty (experimental error), of course, but ideally this may be kept within manageable proportions. It is this innate element of uncertainty that makes bioassay a candidate for statistical consideration.

2. On the basis of response: On the basis of the response evoked in the test subjects of choice, most bioassays may be categorized into one of the following types:

Direct assays: In these the response in the individual test subject is absolute (live, die; response, non-response; etc.) and critical (thresh-hold) levels of the assayed material are determinate, at least within reasonable limits. Computations mainly involve calculation of means and ratios, and estimation of standard errors or confidence limits of such statistics. Example: the cat assay of digitalis.

Graded response-parallel line assays: In these, the response in the individual is proportional to the dose of test substance administered and the degree of response is experimentally determinable. Typically, the degree of response is a linear function of log-dose and the dosage-response regression lines of "Unknown" and "Standard" will be parallel denoting identity or similarity of action. Statistical analysis involves mainly regres-

sion analysis and analysis of variance. With proper design (balanced or partially balanced factorial assays), analysis can be simplified greatly through the use of coefficients. Example: assay of insulin in the rabbit.

Slope-ratio assays: These include mainly the microbiological assays, a group of rather limited general interest in which the degree of measurable response in the individual probably is absolute, but since masses of test subjects (microorganisms) are dealt with, the total response measured, as density, acid formation, etc., approaches a continuous function. Statistical analysis involves multiple regression and relative potency is estimated from the ratio of the partial regression coefficients. Example: microbiological assay of riboflavin.

Quantal response assays: In these, response in the individual test subject is absolute (frequently, live or die) but the critical dose of test material necessary to evoke the response is not directly determinable. Quantitation is achieved through the use of groups of test subjects and determination of the proportion responding to various dosage levels of "Unknown" and "Standard" test products. Following suitable transformation of the data (probits, angles, etc.,) response typically is a linear function of log dose and statistical analysis is essentially similar to that employed with the graded response-parallel line bioassays. Examples: mouse-protective potency assays of typhoid, pertussis and rabies vaccines.

III. REQUIREMENTS OF A VALID BIOASSAY. The following requirements of a "valid" bioassay have evolved from recommendations originally made by Gaddum (1) with modifications made by Bliss, Finney, and others, and are practically universally accepted by students of bioassay. Perhaps the word "valid" should be replaced by "good" or "acceptable."

1. The assay should involve a direct comparison of an unknown with a standard in identical, concomitant tests.

   a. Ideally, the two products should be of essentially equal potency.

2. There should be a significant progressive relationship between dosage and response.

   a. Linear following transformation as required.

   b. Highly significant slope.

   c. No significant curvature; combined or opposed.

3. Dosage-response regression lines for the two products should be parallel, denoting identity or similarity of action.

4. There should be internal evidence of homogeneity (of the data) establishing validity of statistical analysis and adequacy of the testing situation.

5. Analysis should include an estimate of assay error (uncertainty) calculated directly from the data.

Obviously, not all requirements can be applied to each type of assay. Requirements pertaining to slope do not apply to direct assays; those pertaining to parallelism do not apply to slope-ratio assays, etc. However, all do apply to parallel-line graded response assays and most quantal response assays are of similar design.

IV. <u>REDUCTION OF UNCERTAINTY (ERROR) OF BIOASSAYS.</u> All experienced bioassayists are aware of the innate uncertainty and poor reproducibility of such assays as a whole. The degree of variability differs markedly with various assays, perhaps being least with slope-ratio assays and greatest with quantal response assays. This variability can be reduced to some extent in a variety of ways including:

1. Perfection of technique: equipment, reagents, etc.

2. Control of environment: constant temperature, humidity, etc.

3. Increased homogeneity of test subjects: selection of strains, sex and size of test animals; use of litter mates, etc.

4. Use of restricted designs: randomized blocks (complete or incomplete), Latin squares, cross-over designs, confounding, etc.

5. Statistical adjustment of data: covariance analysis, adjusting response data on the basis of a pertinent associated measurement.

6. Increasing the number of observations (test subjects), either by using more subjects per assay or, preferably, by independent replication of the assay as a whole.

In many quantal response assays, particularly assays of vaccines, antisera, etc., most of the above conventional approaches accomplish only modest reduction in assay error. Slopes of the dosage-response regression lines characteristically are low, constituting a major source of assay error, and the main direct compensating approach is to increase the number of test subjects.* A major reduction in assay error, however, would require impractically large numbers of subjects. Practical solution to many of these problems probably lies in the development of assay procedures involving new experimental approaches. If some meaningful response or attribute of the individual test subject can be measured as a continuous variable, a graded response-parallel line assay procedure should be possible. Typically, errors of these assays are much less than of quantal response assays. In some situations, "time to death" has shown promise as a meaningful quantitative response metameter.

V. REQUIREMENTS OF AN ADEQUATE STATISTICAL ANALYSIS.

1. The analysis should provide for the acceptance or rejection of the assay results as a whole; - such acceptability based

*In simplified probit analysis, a crude approximation of the standard error of M (log-ratio of potency) is given by

$$s_M = \frac{1}{b_c} \sqrt{\frac{2}{N_U} + \frac{2}{N_S}}$$

where b_c = combined (average) slope, and N_U and N_S = the number of test subjects assigned to the unknown and standard, respectively

upon the requirements outlined in part III.

2. The analysis should provide for a reliable, unbiased estimate of relative potency that is independent of dosage throughout the maximum possible range.

3. The analysis should provide for an estimate of assay uncertainty, - preferably expressed as confidence limits of the relative potency, - provided meaningful alternatives for action based upon such resultant estimates can be established.

Of the above requirements, the first is considered by this writer to be the most essential and the one most commonly unrecognized or neglected in routine analysis of bioassay data. Specific computational procedures and illustrative examples for all the main types of bioassays are given in standard reference books such as Burn (2), Bliss (3), and Finney (4, 5).

VI. STATISTICAL ANALYSIS OF QUANTAL RESPONSE BIOASSAY DATA. A surprising number and variety of computational procedures for analysis of quantal response bioassay data have been proposed. In terms of statistical rigor and sophistication, they range from simple "quick-and-dirty" graphic approximations to formal iterative procedures involving a degree of complexity and tedious computational detail which is difficult to justify except, possibly, in the most critically extenuating circumstance.

Most, or perhaps all, of these methods have some advantages or disadvantages dependent upon their contemplated use but any critical comparison is far beyond the scope of this presentation. It is consoling to find, however, that they all lead to closely similar estimates of relative potency (or end-points) when applied to truly good data as defined in Part III. Unfortunately, the simpler approximate methods generally do not provide a basis for discrimination between acceptable and non-acceptable data and when applied unwittingly to truly unreliable data may yield estimates which are seriously misleading.

The more commonly used computational procedures can be classified into four general categories. These general categories, examples of methods included in each, and minimal comments regarding each, are given below:

Class	Examples	Comments
Graphic approximations	Miller-Tainter (6).	Minimal calculations; adequate reliability provided good data; some discriminatory power by inspection.
Calculated approximations	Reed-Muench Behrens (7).	Most widely used and probably least reliable of all methods; limited to estimating 50 % endpoint.
Formal procedures	Probit analysis; Bliss (8), Finney (5). Knudsen-Curtis (9).	Laborious calculations; maximum reliability and discriminatory power.
Compromise methods	Litchfield-Wilcoxon (10).	Generally adequate reliability and discriminatory power; appreciably less calculations than formal methods.

Another method, involving a factorial χ^2 approximation, is proposed by this writer. This should be considered a compromise method and is presented in some detail in part VII of this presentation.

The factorial χ^2 approximation is based essentially on analysis of variance of quantal response data espressed in terms of per cent response and log dose. When used with data from balanced factorial bioassays involving a constant number of test subjects per experimental unit, adequate tests for acceptability of the data, the relative potency estimate and an approximation to confidence limits of the relative potency estimate can be obtained with only moderately extensive calculations. Analysis of the data from numerous factorial quantal response bioassays by this method has yielded results in close agreement with those obtained by formal probit analysis (5) and the Knudsen-Curtis method (9).

VII. FACTORIAL χ^2 ANALYSIS OF QUANTAL RESPONSE BIOASSAY DATA. In a previous report (11) the essential computational details of factorial analysis of attribute (enumeration) data, as developed by Brandt, were presented together with illustrations of applications of the method to selected experiments in industrial chemistry. Two forms of the basic formula were presented. The first "(Formula 1)" being the form for

calculating values of χ^2 for individual degrees of freedom from complete factorial experiments in which the experimental units are of equal size, was given as

$$\chi^2_{[1]} = \frac{N^2}{S \times F} \times \frac{T^2}{D}$$

where N = total individuals or observations; S = total successes; F = total failures; T = the total of the sums of products of factorial coefficients and the number of successes in the corresponding experimental units; D = the product of the sums of the squares of the factorial coefficients and the number of individuals per experimental unit; and, the subscript in brackets indicates the degrees of freedom. Either of the outcomes (yes or no, response or non-response, survival or death, etc.) can be designated as success; the other outcome as failure.

In many instances, quantal response bioassay data can be subjected to factorial χ^2 analysis:, the major restrictions being that the experimental units are of equal size and that successive doses of the independent variable (i. e., the toxic or protective substance being assayed) differ by a constant interval when expressed in appropriate units of measurement. In most (perhaps all) assays of immunologically active substances, the successive doses (levels of X) should be increased or decreased in a geometric series such as 1, 2, 4, 8, 16; 1, 3, 9, 27:, etc., as the differences between the logarithms of successive doses are constant in value. When these restrictions are complied with, factorial coefficients (3) can be used directly in analysis of the data and χ^2 values can be computed by the formula given above. In this manner it is possible to obtain statistical information regarding the validity or adequacy of the data (Part III) and, as shown below, to obtain a direct estimate of relative potency and its approximate confidence limits.

The procedures are illustrated with actual examples of both 2-dose (4-point) and 3-dose (6-point) assays of the mouse protective potency of typhoid vaccine performed by the author at the Army Medical Service Graduate School.* Details of the assay procedure employed have been published previously (12); attention here will be limited primarily to statistical treatment of the data.

---

* Now known as Walter Reed Army Institute of Research.

1. Factorial χ^2 analysis of a 2-dose quantal response bioassay

As a part of a study to determine the reproducability of mouse protection potency assays of typhoid vaccine (13), a series of 6 assays were run on identical aliquots of a reference vaccine. The aliquotes were identified only as A and B and prior to the assay it was decided to calculate their relative potency, B as per cent of A. Data from the sixth trial are reproduced in Table I.

Table I
Two-dose assay of the mouse protective potency of typhoid vaccines
(Survivals/totals)

Vaccine	Vaccine dose (ml)	
	0.015	0.15
A	5/20	13/20
B	2/20	15/20

For factorial χ^2 analysis, these data are rearranged to the form given in Table IA. For purposes of obtaining tests of significance (χ^2) it is of no consequence in which order the vaccines are entered in the table or which comparison groups are assigned + and - coefficients. However, in the estimation of relative potency, slope, etc., computations are more convenient if certain orders are followed. For the comparison between products (designated as comparison a), positive coefficients should be assigned to the "unknown" (vaccine B in this case). Likewise, for the estimation of slope (comparison a), positive coefficients should be assigned to the higher dose level. Assignment of coefficients to the interaction comparison (ab) is uniquely determined as the cross products of coefficients for the first 2 comparisons, of course. This assignment of coefficients is consistent with that employed by Bliss (3) and others.

Table IA
Factorial χ^2 analysis of the data on Table I

Vaccine	B (unknown)		A (standard)								
Dose	Low	High	Low	High							
Success (survivors) / 20	2	15	5	13	Σ+	Σ-	T	T^2	D	T^2/D	χ^2*
Comparisons											
a Unknown vs standard	+	+	-	-	17	18	-1	1	80	0.0125	0.05
b Slope (high vs low dose)	-	+	-	+	28	7	21	441	80	5.5125	22.38
ab Departure from parallelism (products x doses)	-	+	+	-	20	15	5	25	80	0.3125	1.27

$$^{*}\chi^2 = \frac{N^2}{S \times F} \times \frac{T^2}{D} = \frac{80^2}{35 \times 45} \times \frac{T^2}{D} = 4.06 \times \frac{T^2}{D}.$$

Evidence of assay validity: All calculations are performed in the manner previously described (11). From comparison a, it is found that the 2 vaccines do not differ appreciably in total effect ($\chi^2[1] = 0.05$). From comparison b it can be seen that there is a highly significant relationship between dosage and response ($\chi^2[1] = 22.38$), and by comparison ab it is determined that there is no significant departure from parallelism exhibited by the dosage response lines for the unknown and standard. No information is available concerning curvature of the dosage response curves. Such can be obtained only when 3 or more dosage levels are employed.

As the assay actually was conducted, the 20 mice in each experimental unit were not handled as a single group but as 4 independent groups of 5 each. These groups were selected, assigned spaces in the test room, immunized and challenged in random order and the number of survivors originally were recorded per group of 5. Thus it is possible to calculate a "within groups" χ^2 with 12 degrees of freedom which can be used as a measure of internal homogeneity (requirement 4). The procedure will be illustrated with data from the next example (Table II).

Estimation of relative potency: It is possible to obtain an estimate of relative potency (RP) from the data and calculations of Table IA by use of the formula for estimating relative potency from

a 2-dose factorial assay as given by Bliss (3).

$$M = \frac{i \times T_a}{T_b}$$

where M = the log ratio of potency; i = the log-dose increment*; and, T_a and T_b are the values in the column headed T for comparisons a and b, respectively. In this assay, the dosage increment was 10-fold, so i = log 10 = 1. $T_a = -1$ and $T_b = 21$. Substituting these values in the formula, M is calculated as

$$M = \frac{1 \times -1}{21} = -0.0476.$$

This value is a logarithm and must be converted to the usual form $\bar{1}.9524$. The antilogarithm of $\bar{1}.9524$ is the relative potency which is found to be 0.896; or, in terms of percentage, vaccine B is 89.6 per cent as potent as vaccine A. This estimate is in reasonably close agreement with that obtained by probit analysis, 85.4 per cent.

Approximate confidence limits of relative potency: It also is possible to obtain an approximation of the confidence limits of the relative potency estimate from the data and calculations presented in Table IA. This is most easily done by first determining the approximate confidence interval for $M(CI'_M)$ which for a 2-dose assay is calculated as

$$CI'_M = \frac{1.96 \times 2n/\sqrt{N} \times i}{T_b} \; **$$

where n = individuals per experimental unit; N = 4n or grand total individuals, and i and T_b have the same meaning as before. The

*Logarithms of dosage increments from 2-fold to 10-fold are tabulated in Table 1, Appendix 1 and designated as constants $c_{M.2}$.

**The term "confidence interval" typically is used to denote the entire range included between lower and upper confidence limits. The quantity approximated by CI'_M, as used here, is one-half the entire range expressed in logarithmic units. Derivation of this approximation is given in Appendix II to this paper.

95 per cent confidence limits of M then are determined as

$$M \pm CI'_M$$

and the 95 per cent confidence limits of the relative potency (95% CL_{RP}) are found as the antilogarithms of these 2 values.

$$95\%\ CL_{RP} = \text{antilogarithms of } M - CI'_M \text{ and } M + CI'_M.$$

These limits will be in the form of ratios which can be converted to percentage by multiplying by 100. For the illustrative problem dealt with here (Tables I and IA)

$$CI'_M = \frac{1.96 \times 40/\sqrt{80} \times \log 10}{21} = 0.4174$$

Then

$$95\%\ CL_M = -0.0476 + 0.4174 = -0.4650 \text{ and } 0.3698$$

$$= \bar{1}.5350 \text{ and } 0.3698$$

Taking antilogarithms

$$95\%\ CL_{RP} = 0.34 \text{ and } 2.34$$

or 34% and 234%.

Thus, the best estimate of relative potency (B as per cent of A) is 89.6 per cent and the odds are approximately 19 out of 20 that the true potency is between 34 and 234 per cent.

For a factorial assay of set design, where i and n are constant, assay to assay, the foregoing calculations can be simplified as all elements in the formula for CI'_M will be the same

except for T_b. Thus constants for 2-dose assays ($C_{I.2}$) and 3-dose assays ($C_{I.3}$) for fold-increments of dosage from 2 to 10, and for values of n from 10 to 20, have been calculated and are presented in Appendix I, Tables 2 and 3.

It must be emphasized that this estimate of the confidence limits of the relative potency is only an approximation. Yet the results obtained were in reasonably close agreement with those obtained by probit analysis, 31.6 and 230.4 per cent.

2. Factorial χ^2 analysis of a 3-dose quantal response bioassay.

Factorial χ^2 analysis of a 3-dose quantal response assay for determining the validity of the assay and the estimation of relative potency and approximate 95% confidence limits of the potency estimate, are illustrated with data from another typhoid vaccine mouse protection potency test performed at the Army Medical Service Graduate School. The vaccines tested were a routine production lot (unknown) and a reference standard. Results of the assay are summarized in Table II, and are arranged in the form suitable for factorial χ^2 analysis in Table IIA.

Table II

Three-dose assay of the mouse protective potency of an unknown typhoid vaccine in respect to a standard

(Survivors/totals)

Vaccine	Vaccine dose (ml)		
	0.02	0.08	0.32
Unknown	1/10	5/10	8/10
	1/10	7/10	9/10
Standard	2/10	4/10	8/10
	1/10	5/10	7/10

Table IIA
Factorial χ^2 analysis of the data of Table II

Vaccine	Unknown			Standard									
Dose	D_1	D_2	D_3	D_1	D_2	D_3							
Success (survivors)	1	5	8	2	4	8							
10	1	7	9	1	5	7							
Successes/20	2	12	17	3	9	15	$\Sigma+$	$\Sigma-$	T	T^2	D	T^2/D	χ^{2*}
Comparisons													
a Unknown vs standard	+	+	+	-	-	-	31	27	4	16	120	0.13	0.52
b Slope	-	0	+	-	0	+	32	5	27	729	80	9.11	36.44
ab Parallelism	-	0	+	+	0	-	20	17	3	9	80	0.11	0.44
c Combined curvature	+	-2	+	+	-2	+	37	42	5	25	240	0.10	0.40
ac Opposed curvature	+	-2	+	-	+2	-	37	42	5	25	240	0.10	0.40

$$^{*}\chi^2 = \frac{N^2}{S \times F} \times \frac{T^2}{D} = \frac{120^2}{58 \times 62} \times \frac{T^2}{D} = 4.00 \times \frac{T^2}{D}$$

Between groups within experimental units:

$$\chi^2_{\boxed{6}} = 4.00 \times \frac{(1-1)^2 + (7-5)^2 + (9-8)^2 + (2-1)^2 + (5-4)^2 + (8-7)^2}{20} = 4.00 \times \frac{8}{20}$$

$$= 1.60.$$

There is little need for comment regarding the computational procedure employed. Factorial coefficients were assigned in conventional order (3) and χ^2 values for each comparison were computed in the manner previously described. Calculation of χ^2 "between groups within experimental units" was accomplished by summation of all T^2/D values between pairs of groups of 10 each and multiplying the total by the constant $\frac{N^2}{S \times F}$.

Evidence of validity

There was no evidence of significant differences between the pairs of groups within experimental units ($\chi^2_{\boxed{6}} = 1.60$). This yields assurance that the randomization procedures employed during the assay were adequate to prevent appreciable bias due to technical and environmental factors. Since 3 dosage levels of vaccine were employed, it was possible to gain information regarding curva-

ture of the dosage response lines, both combined and in opposition. There was no evidence of systematic departure from linearity. Thus, all requirements for assay validity (Part III) were satisfied.

Estimation of relative potency

The relative potency of the unknown in respect to the standard was estimated by the formula given by Bliss (3) for calculating M in 3-dose factorial assays

$$M = \frac{4 \times i \times T_a}{3 \times T_b} \quad *$$

The dosage increment employed in this assay was 4-fold, so i= log 4 = 0.6021. Substituting calculated values of T_a and T_b into the formula, M was calculated as

$$M = \frac{4 \times 0.6021 \times 4}{3 \times 27} = 0.1189$$

and the relative potency = 100 x antilog 0.1189 = 131.5 per cent.

Approximate confidence limits of relative potency

The formula for estimating the approximate confidence interval of M in a 3-dose assay differs from that for 2-dose assay only in that 4n must be substituted for 2n. Thus, for a 3-dose factorial assay.

$$CI'_M = \frac{1,96 \times 4n/\sqrt{N} \times i}{T_b}$$

* Values of $\frac{4 \times i}{3}$ dosage increments of 2-fold through 10-fold have been calculated and are given as constants $c_{M.3}$ in Table 1, Appendix I. M is determined by multiplying the ratio T_a/T_b by the appropriate value of $c_{M.3}$ (0.8020 in this example).

For the data dealt with here (Tables II and IIA), $n = 20$, $N = 120$, and $i = \log 4 = 0.6021$. Then

$$CI'_M = \frac{1.96 \times 80/\sqrt{120} \times 0.6021}{27}*$$
$$= 0.3192.$$

The confidence limits of M are found as

$$95\,\%\ CL_M = M \pm CI'_M$$

$$= 0.1189 \pm 0.3192 = -0.2003 \text{ and } 0.4381$$

$$= \overline{1}.7997 \text{ and } 0.4381$$

Then the 95 per cent confidence limits of the relative potency are obtained as the antilogarithms of these values.

$$95\%\ CL_{RP} = 0.63 \text{ and } 2.74$$

or

$$= 63 \text{ and } 274 \text{ per cent}$$

These data also were analyzed by the probit analysis. The relative potency estimate was 132.2 per cent and the 95 per cent confidence limits were 64.2 per cent and 272.2 per cent.

3. Resumé of computational procedure: Chi square analysis of quantal response factorial assays yielding (1) statistical evidence regarding reliability of the data, (2) an estimate of relative potency, and (3) approximate confidence limits of the relative potency, involves a series of 7 main steps.

1. Arrange the data on a work sheet of the form used in Tables IA and IIA.

2. Assign the factorial coefficients in accordance with the actual design of the experiment. Compute N^2/SxF from the grand

---

*Constants $c_{I.3}$ for estimating values of CI'_M in 3-dose factorial assays for dosage increments of 2-fold through 10- fold and for values of n from 10 through 20, have been calculated and are given in Table 3 of the appendix. For this problem, $c_{I.3} = 8.6183$. This divided by 27 $(T_b) = 0.3192$, the same as calculated above.

totals and then $\Sigma+$, $\Sigma-$, T, T^2, D, T^2/D and χ^2 for each comparison (row). Also, if data on subgroups within experimental units are available, calculate the "between groups" χ^2 (cf. Table IIA). From the various values of χ^2 determine if there is sufficient evidence of validity to justify estimation of potency.

3. If justified, compute the ratio T_a/T_b and calculate M as:

   a. Two dose assay: $M = i \times T_a/T_b$. Values of i are given as the constants $c_{M.2}$ in Table 1, Appendix I.

   b. Three-dose assay: $M = \frac{4 \times i}{3} \times T_a/T_b$. Values of $\frac{4 \times i}{3}$ are given as the constants $c_{M.3}$ in Table 1, Appendix I.

4. Determine the relative potency (RP) as a ratio or percentage as antilog M, or as 100 x antilog M, respectively.

5. Compute CI'_M as:

   a. Two-dose assay:

$$CI'_M = \frac{1.96 \times 2n\sqrt{N} \times i}{T_b}$$

   or, using constants $c_{I.2}$ from Table 2, Appendix I:

$$CI'_M = \frac{c_{I.2}}{T_b}.$$

   b. Three-dose assay:

$$CI'_M = \frac{1.96 \times 4n/\sqrt{N} \times i}{T_b}$$

   or, using constants $c_{I.3}$ from Table 3, Appendix I:

$$CI'_M = \frac{c_{I.3}}{T_b}.$$

6. Calculate the 95 per cent conficence limits of M as

$$95\%\ CL_M = M \pm CI'_M.$$

7. Determine the 95 per cent confidence limits of the relative potency as

$$95\%\ CL_{RP} = \text{antilog } M - CI'_M \text{ and antilog } M + CI'_M.$$

If it is desired to express the limits as percentages, multiply each value by 100.

APPENDIX I

Table 1

Values of $c_{M.2}$ and $c_{M.3}$ for obtaining estimates of M, the log ratio of potency, from 2-dose and 3-dose factorial assays

$$(M = c_{M.i} \times T_a/T_b)\ *$$

Fold-increment in dosage	$c_{M.2}$ (2-dose assays)	$c_{M.3}$ (3-dose assays)
2	0.3010	0.4013
3	0.4771	0.6361
4	0.6021	0.8028
5	0.6990	0.9320
6	0.7782	1.0376
7	0.8451	1.1268
8	0.9031	1.2041
9	0.9542	1.2722
10	1.0000	1.3333

*Relative potency = antilog M.

Relative potency in % = 100 x antilog M.

This Page Intentionally
Left Blank

APPENDIX I
Table 2
Values of $c_{I.2}$ *

Constants for calculating CI'_M, the confidence interval of M, for 2-dose factorial assays. Select the values of $c_{I.2}$ determined by the dosage increment (row) and group size (column).

$$\left(CI'_M = \frac{c_{I.2}}{T_b}\right)$$

Fold-increment in dosage	n (individuals per group)										
	10	11	12	13	14	15	16	17	18	19	20
2	1.8656	1.9567	2.0437	2.1271	2.2074	2.2849	2.3598	2.4325	2.5030	2.5716	2.6384
3	2.9571	3.1014	3.2393	3.3716	3.4989	3.6217	3.7405	3.8556	3.9674	4.0761	4.1820
4	3.7318	3.9104	4.0880	4.2550	4.4156	4.5705	4.8658	4.7205	5.0068	5.1440	5.2776
5	4.3324	4.5439	4.7459	4.9398	5.1263	5.3061	5.4802	5.6488	5.8126	5.9718	6.1270
6	4.8233	5.0588	5.2837	5.4995	5.7071	5.9073	6.1011	6.2889	6.4712	6.6485	6.8212
7	5.2379	5.4937	5.7379	5.9722	6.1977	6.4152	6.6256	6.8295	7.0275	7.2200	7.4076
8	5.5974	5.8707	6.1317	6.3821	6.6231	6.8554	7.0803	7.2982	7.5098	7.7155	7.9160
9	5.9141	6.2029	6.4786	6.7432	6.9978	7.2433	7.4809	7.7112	7.9347	8.1521	8.3639
10	6.1980	6.5006	6.7896	7.0669	7.3337	7.5910	7.8400	8.0813	8.3156	8.5434	8.7654

*Calculated as $\frac{1.96 \times 2n \times i}{N}$

This Page Intentionally
Left Blank

APPENDIX I

Table 3

Values of $c_{I.3}$ *

Constants for calculating CI'_M, the confidence interval of M, for 3-dose factorial assays.

Select the values of $c_{I.3}$ determined by the dosage increment (row) and group size (column).

$$(CI'_M = \frac{c_{I.3}}{T_b})$$

Fold-increment in dosage	n (individuals per group)										
	10	11	12	13	14	15	16	17	18	19	20
2	3.0465	3.1952	3.3373	3.4736	3.6047	3.7313	3.8536	3.9722	4.0874	4.2013	4.3085
3	4.8289	5.0646	5.2898	5.5058	5.7136	5.9142	6.1081	6.2961	6.4787	6.6593	6.8291
4	6.0941	6.3915	6.6757	6.9483	7.2106	7.4638	7.7084	7.9457	8.1761	8.4041	8.6183
5	7.0749	7.4202	7.7501	8.0665	8.3710	8.6649	8.9490	9.2245	9.4919	9.7566	10.0053
6	7.8765	8.2609	8.6282	8.9805	9.3195	9.6467	9.9630	10.2697	10.5674	10.8620	11.1390
7	8.5536	8.9711	9.3670	9.7525	10.1207	10.4760	10.8195	11.1525	11.4759	11.7958	12.0966
8	9.1406	9.5868	10.0130	10.4219	10.8153	11.1950	11.5620	11.9179	12.2635	12.6054	12.9268
9	9.6578	10.1292	10.5796	11.0116	11.4272	11.8285	12.2162	12.5923	12.9574	13.3186	13.6582
10	10.1214	10.6154	11.0874	11.5401	11.9757	12.3962	12.8026	13.1967	13.5793	13.9579	14.3138

* Calculated as $\frac{1.96 \times 4n \times i}{N}$

This Page Intentionally
Left Blank

APPENDIX II

Approximation of the Confidence Interval of M

The standard error of M (s_M) from a balanced factorial bioassay is given by Bliss (3) as:

$$s_M = \frac{s}{b_c}\sqrt{\frac{4}{N}\left[1 + \frac{D^2}{B^2 - s^2t^2}\right]}$$

where D^2 = mean square between products; B^2 = mean square for combined slope; s^2 = mean square; t = Student's statistic; N = total number of test subjects (possible responses); and b_c is the combined or average slope of the dose-response regression line.

For a 2-dose (4-point) assay, b_c is estimated as $\frac{Tb}{2 \times i \times n}$; for a 3-dose (6-point) assay, as $\frac{Tb}{4 \times i \times n}$. In these, Tb is found as shown in Tables IA and IIA, i is the log ratio of dosage increment, and n is the number of test subjects per experimental group.

In a good bioassay (statistically acceptable), D^2 will be small and B^2 will be large. Thus, the quantity enclosed in brackets approaches unity and can be ignored. In the binomial, the variance (s^2) has a maximum value of 0.25 and s has a maximum value of 0.5. In a balanced assay of fixed design, N will be 4n or 6n for a 2-dose and 3-dose assay, respectively. Substituting the appropriate formula for b_c as given above, and introducing $t_\infty = 1.96$, the confidence intervals of M can be reduced to the following approximations:

2-dose assay: $$CI_M = \frac{(1.96 \times 2n \times i)/\sqrt{N}}{T_b}$$

3-dose assay: $$CI_M = \frac{(1.96 \times 4n \times i)/\sqrt{N}}{T_b}$$

These approximations were used for calculating the constants presented in Tables 2 and 3 of Appendix I.

REFERENCES

1. Emmens, C. W., Principles of Biological Assay, London, Chapman and Hall, Ltd., 1948.

2. Burn, J. H., Finney, D. J., and Goodwin, L. G., Biological Standardization. Oxford Univ. Press, New York, ed. 2, 1950.

3. Bliss, C. I., The Statistical Method in Biological Assay. Hafner Publ. Co., New York, 1952.

4. Finney, D. J., Statistical Method in Biological Assay. Hafner Publ. Co., New York, 1952.

5. Finney, D. J., Probit Analysis. Cambridge Univ. Press, London, 1947.

6. Miller, L. C., and Tainter, M. L., Estimation of the $E.D._{50}$ and its error by means of logarithmic - probit graph paper. Proc. Soc. Exp. Biol. and Med., 57: 261-264, 1944.

7. Reed, L. J., and Muench, H., A simple method of estimating fifty per cent endpoints. Am. J. Hygiene, 27: 493-497, 1938.

8. Bliss, C. I., The determination of dosage-mortality curves from small numbers. Quart. J. Pharm., 11: 192-216, 1938.

9. Knudsen, L. F., and Curtis, J. M., The use of the angular transformation in biological assay. J. Am. Stat. Assoc., 42: 282-296, 1947.

10. Litchfield, J. T., and Wilcoxon, F., A simplified method of evaluating dose-effect experiments. J. of Pharm. and Expt. Therap., 96: 99-113, 1949.

11. Batson, H. C., Applications of factorial χ^2 analysis to experiments in chemistry. Transactions, 10th Annual Meeting, ASQC, p. 9-23, 1956.

12. Batson, H. C., Brown, M., and Oberstein, M., The mouse-protective potency assay of typhoid vaccine as performed at the Army Medical Service Graduate School. Pub. Health Rep. 66: 789-806, 1951.

13. Batson, H. C., Brown, M., and Oberstein, M., An adaptation of quality control chart methods to bacterial vaccine potency testing. J. Bact. 61: 407-419, 1951.

This Page Intentionally
Left Blank

ATTENDEES TO THE EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

24-26 Oct. 1962

Abelson, Robert P.	Yale University, New Haven, Conn.
Addelman, Sidney	Research Triangle Institute, Durham, N.C.
Altman, Peter S.	U. S. Naval Weapons Lab, Dahlgren, Va.
Atkinson, John C.	Army Chem. Center, Md.
Atwood, Earl	WRAIR, Washington, D. C.
Baird, H. Robert	National Institutes of Health, (NIH) Bethesda, Md.
Bales, Robert A.	Tech. Oper., Inc. Ft. Belvoir, Va.
Banton, Clarence W.	Army Tank Automotive Center, Detroit Arsenal
Barker, Dana W.	Combat Dev. Comd., Ft. Belvoir, Va.
Barker, Lewellys	NIH, Bethesda, Md.
Batson, Herbert C.	University of Illinois, Urbana, Ill.
Beckmann, Edward M.	Rock Island Arsenal, Rock Island, Ill.
Benson, Jacob	Ft. Monmouth, New Jersey
Benson, Karl L.	Office of the Surgeon General, Wash., D.C.
Bombara, Elwood L.	Redstone Arsenal, Huntsville, Ala.
Boyd, Donald M.	Res. Analysis Corp., Bethesda, Md.
Boyle, Douglas G.	Hercules Powder Co., Magna, Utah
Brown, Ralph E.	Frankford Arsenal, Philadelphia, Pa.
Brown, William A.	Dugway Proving Ground, Dugway, Utah
Bruno, O. P.	Ballistic Res. Labs, Aberdeen Proving Ground, Maryland
Bulfinch, Alonzo	Picatinny Arsenal, Dover, N. J.
Burdick, Donald S.	Duke University, Durham, N. C.
Cabaniss, Jelks H., Jr.	Department of Math, U. S. Military Academy, West Point, N. Y.
Carter, Frederick L., Jr.	USA Chem. Labs., Ft. Detrick, Md.
Chernoff, Herman	Stanford University, Stanford Calif.
Coberly, Robert H.	Rock Island Arsenal, Rock Island, Ill.
Cohen, A. Clifford	University of Georgia, Athens, Ga.
Collins, Elmer F.	Naval Receiving Station Annex
Cook, Kenneth G.	Applied Psychology Corporation, Arlington, Va.

Cornfield, Jerome	NIH, Bethesda, Md.
Cox, Claire B.	NIH, Bethesda, Md.
Cox, Edwin L.	Agr. Res. Service, Beltsville, Md.
Cox, Paul C.	White Sands Missile Range, N. M.
Craw, Alexander R.	USA Bio. Labs., Ft. Detrick, Md.
Creecy, Donald B.	National Security Agency, Ft. Meade, Md.
Culbertson, George E.	Hercules Powder Co., Radford Arsenal, Va.
Daggit, Edward A.	U. S. Military Academy, West Point, N. Y.
DeArmon, Ira A.	Edgewood Arsenal, Md.
Dobrindt, Gerard T.	Aberdeen Proving Ground, Md.
Dowling, Jasper B.	Picatinny Arsenal, Dover, N. J.
Dressel, Francis D.	Army Research Office, Durham, N. C.
Dorn, H. F.	NIH, Bethesda, Md.
Eckard, Cecil O.	Deseret Test Center, Ft. Douglas, Utah
Einbinder, Seymour K.	Picatinny Arsenal, Dover, N. J.
Eisenhart, Churchill	National Bureau of Standards, Washington, D. C.
Eisler, Stanley L.	Rock Island Arsenal, Ill.
Eissner, Robert M.	Aberdeen Proving Grounds, Md.
Ellner, Henry	Army Chemical Center, Md.
Ewart, Wade H.	Redstone Arsenal, Huntsville, Ala.
Federer, Walter T.	Math Res. Center, University of Wisconsin, Madison, Wisconsin
Fiess, Edward R.	U.S. Army Electronic R&D Lab., Ft. Monmouth, N. J.
Finkel, Peter W.	Res. Analysis Corp., Bethesda, Md.
Foster, Walter D.	Chemical, Biological, & Radiological Laboratories, Ft. Detrick, Md.
Frishman, Fred	Army Research Office, Washington 25, D. C.
Frick, James W.	Cornell Aero. Lab., Arlington, Va.
Fulton, Charles L.	Frankford Arsenal, Philadelphia, Pa.
Galbraith, Alan S.	Army Research Office, Durham, N. C.
Gardner, Roberta A.	NIH, Bethesda, Md.
Gauld, Ross L.	Walter Reed Army Institute of Research, Washington, D. C.

Gebhard, R.	Human Resources Research Office Alexandria, Va.
Geisser, Seymour	NIH, Bethesda, Md.
Grant, Roy L.	US Bureau of Mines, Pittsburgh, Pa.
Greenberg, Bernard G.	University of North Carolina Chapel Hill, N. C.
Greenhouse, Samuel W.	NIH, Bethesda, Md.
Greenstone, Reynold	Operations Res., Inc., Silver Springs, Md.
Greenwood, Joseph A.	Armed Forces Radiobiol. Research Institute, Bethesda, Md.
Grubbs, Frank E.	Aberdeen Proving Ground, Md.
Haines, Bertram W.	Ft. Detrick, Md.
Hamilton, Eugene L.	Office of the Surgeon General, Washington, D. C.
Hamman, J. Wayne	Armed Forces Food & Cont. Inst., Chicago, Ill.
Harshbarger, Boyd	Virginia Polytechnic Institute, Blacksburg, Va.
Hartley, Herman O.	Iowa State University, Ames, Ia.
Heermann, Emil F.	U. S. Army Personnel Research Office, Washington, D. C.
Heifetz, Samuel	Aberdeen Proving Ground, Md.
Helmbold, Robert L.	Burlington, Mass.
Hill, Floyd I.	Cornell Aero. Lab., Arlington, Va.
Hixson, Eugene E.	National Aeronautics & Space Administration, Greenbelt, Md.
Hoppes, Harrison N.	Res. Analysis Corp., Bethesda, Md.
Horner, Theodore W.	Booz-Allen Appl. Research, Bethesda, Md.
Hundley, Douglas V.	Army Chemical Center, Md.
Hutton, Matthew I.	Edgewood Arsenal, Md.
Inselmann, Edmund H.	Frankford Arsenal, Philadelphia, Pa.
Isaac, Gerhard J.	Med. Res. & Nutr. Lab., Denver, Colo.
Jebe, Emil H.	University of Michigan, Ann Arbor, Mich.
Jessup, Gordon L., Jr.	USA Bio. Labs., Ft. Detrick, Md.
Johnson, Cecil D.	USA Personnel Research Office, Washington, D. C.

Johnson, Dennis E.	Arthur D. Little, Inc., Cambridge, Mass.
Johnson, Wilbur V.	Res. Analysis Corp., Bethesda, Md.
Karp, Abraham E.	Department of Army, Bethesda, Md.
King, Donald W.	Westat Research Analysts
Kirby, William H. Jr.	Aberdeen Proving Ground, Md.
Kornack, Wallace R.	Atomic Energy Commission, Washington, D. C.
Krane, Scott A.	C-E-I-R, Inc., Dugway, Utah
Krimmer, Manfred W.	Picatinny Arsenal, Dover, N. J.
Ku, Hsien H.	National Bureau of Standards, Washington, D. C.
Kupperman, Morton	National Space Agency, Ft. Meade, Md.
Kurkjian, Badrig M.	Harry Diamond Laboratories, Washington, D. C.
Langlie, H. Joseph	Ford Aeronutronic, Newport Beach, Calif.
Lavin, George I.	Aberdeen Proving Ground, Md.
LeMaster, Vaughn	Ammo. Proc. & Supply, Joliet, Ill.
Levin, Sheldon G.	Harry Diamond Laboratories, Washington, D. C.
Littrell, Robert	Res. Analysis Corp., Bethesda, Md.
Love, Benjamin T.	RCA Service Co., Tuscon, Ariz.
Lucas, Henry L.	North Carolina State College, Raleigh, N. C.
Maisel, Herbert	U. S. Army Strategy & Tactics Analysis Group, Washington, D. C.
Mall, Adolph W.	Frankford Arsenal, Philadelphia, Pa.
Mallios, William S.	Booz-Allen Appl. Research, Inc., Bethesda, Md.
Maloney, Clifford J.	NIH, Bethesda, Md.
Mandelson, Joseph	Army Chemical Center, Md.
Marshall, Jack A.	University of Chicago, Chicago, Ill.
Marthens, Arthur S.	BuShips, Washington, D. C.
Martin, Cyrus	Corps of Engineers, Ft. Belvoir, Va.
Masaitis, Ceslovas	Aberdeen Proving Ground, Md.
Mauss, Bess Day	Washington, D. C.
Metcalfe, Samuel N.	Ft. Detrick, Md.
Mitchell, Thomas D.	Naval Avionics Facility, Indianapolis, Ind.
Mann, G. V.	Vanderbilt University, Nashville, Tenn.

Morrison, Donald F.	NIH, Bethesda, Md.
Mosimann, James E.	NIH, Bethesda, Md.
McIntosh, William B.	U. S. Army Electronic Proving Ground, Ft. Huachuca, Ariz.
Natrella, Mary G., Mrs.	National Bureau of Standards, Washington, D. C.
O'Brien, Richard J.	Rock Island Aresenal, Ill.
Okano, Fred	National Aeronautics & Space Administration, Washington, D. C.
Olkin, Ingram	Stanford University, Stanford, Calif.
O'Malley, Joseph P.	NIH, Bethesda, Md.
Orleans, Beatrice S.	BuShips, Washington, D. C.
Owen, Donald B.	Sandia Corp., Albuquerque, N. M.
Pabst, W. R.	Bureau of Naval Weapons, Washington, D. C.
Panos, Robert V.	C-E-I-R, Inc., Dugway, Utah
Parks, Albert B.	Harry Diamond Laboratories, Washington, D. C.
Parnell, Benjamin L.	Walter Reed Army Medical Center Washington, D. C.
Parrish, Gene B.	Army Research Office, Durham, N. C.
Pearson, Egon S.	London University, London, England
Perry, Virginia W.	QM Research & Engineering, Field Evaluation Agency, Ft. Lee, Va.
Pierce, William L.	Res. Analysis Corp., Bethesda, Md.
Pradko, Fred	Army Tank-Auto Comd., Center Line, Mich.
Prairie, Richard R.	Sandia Corp., Albuquerque, N. M.
Purtell, John P.	Watervliet Arsenal, Watervliet, N. Y.
Raisbeck, Leo R.	Rock Island Arsenal, Ill.
Rhian, Morris A.	Edgewood Arsenal, Md.
Rider, Paul R.	Aero Research Lab., Wright-Patterson AFB
Riggs, Charles W.	USA Bio. Labs., Ft. Detrick, Md.
Romano, Albert	National Bureau of Standards Washington, D. C.
Schneiderman, Marvin A.	National Cancer Institute, Bethesda, Md.
Smith, N. M.	Res. Analysis Corp., Bethesda, Md.

Schnell, Raymond B.	Chemical Res. & Dev. Lab., Edgewood, Md.
Schoolman, Harold M.	Veterans Administration Hospital, Hines, Ill.
Schroeder, Bruno W.	U. S. Army Electronic Proving Ground, Ft. Huachuca, Ariz.
Schwarz, Richard A.	U. S. Atomic Energy Commission, Germantown, Md.
Shaw, Thomas R.	Operations Research, Inc., Silver Spring, Md.
Sheehy, Robert W.	Dept. Biologics Research Walter Reed Army Institute of Research, Washington, D. C.
Sloane, Harry S.	Dugway Proving Ground, Dugway, Utah
Sobelman, Sidney	Picatinny Arsenal, Dover, N. J.
Speckman, Janace A.	Bureau of Standards, Washington, D. C.
Stallmann, Friedemann W.	Veterans Administration, Washington, D. C.
Stearns, Edward R.	Army Chemical Center, Md.
Steck, George P.	Sandia Corporation, N. M.
Steedman, Joseph E.	Aberdeen Proving Ground, Md.
Stephanides, Agatha, Mrs.	U. S. Army Strategy & Tactics Analysis Group, Bethesda, Md.
Taft, Everett	University of Michigan, Ann Arbor, Mich.
Tarlow, Irving	QM Res. & Eng. Cmd., Natick, Mass.
Thompson, W. A.	University of Delaware, Newark, Md.
*Wadley, Francis M.	Chem. Corps Biol. Lab, Ft. Detrick, Md.
Wallach, Harold C.	Census Bureau, Suitland, Md.
Walner, Arthur H.	N. Y. Naval Shipyard, Brooklyn, N. Y.
Weinstein, Joseph	Elect. Res. & Dev. Laboratories, Ft. Monmouth, N. J.
Weintraub, Gertrude	Picatinny Arsenal, Dover, N. J.
White, John E.	Bridgewater College, Bridgewater, Va.
Wilburn, Nicholas T.	Ft. Monmouth, N. J.
Wilks, S. S.	Princeton University, N. J.
Wolman, William	National Aeronautics & Space Administration, Washington, D. C.
Wentworth, Paul	WRAIR, Washington, D. C.
Youden, William J.	National Bureau of Standards, Washington, D. C.
Vivona, Stefano	WRAIR, Washington, D. C.

Zelen, Marvin	Math Research Center, University of Wisconsin, Madison, Wis.
Zundel, Ivan B.	C-E-I-R, Inc., Dugway, Utah